

STIC-Biotech/ChemLib

94991

From: Ibrahim, Medina A.  
Sent: Tuesday, May 27, 2003 7:21 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/977, 137

RECEIVED

MAY 27 2003

STIC

Please search the following:

1. SEQ ID NO:3-4.
2. oligo search of SEQ ID NO:3-4.

Please search both commercial and patent databases (including pending). Thanks

Medina A. Ibrahim  
Patent Examiner  
GAU:1638  
CM1-9E03  
mailbox-9E12  
(703)306-5822

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/28  
Date Completed: 6/7  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 07:34:19 ; Search time 1511 Seconds  
(without alignments)  
3440.603 Million cell updates/sec

Title: US-09-977-137A-3

Perfect score: 321  
Sequence: 1 atgacacactgcgagagagc.....aggggaatgttctctgcgcg 321

Scoring table: OLIGO-NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: \*  
1: em\_estbda: \*  
2: em\_esthum: \*  
3: em\_estlin: \*  
4: em\_estmuy: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hlc: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hlc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estum: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vrt: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rtd: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	5.9	305	14	C91774 C91774 Rice
2	19	5.9	777	13	B1480603 EST0056 H
3	19	5.9	1099	12	BE791590 B1480603 EST0056 H
4	18	5.6	344	9	A1824830 w02d07.x
5	18	5.6	362	10	BE227306 BE227306 894030B09
6	18	5.6	398	10	AW628641 h137n06.x

7	18	5.6	409	12	BF086409	BF086409 RC6-GN007
8	18	5.6	425	13	BM487398	BM487398 psm2n.pk0
9	18	5.6	425	13	BM491850	BM491850 psp2n.pk0
10	18	5.6	452	12	BG713248	BG713248 p911n.pk0
11	18	5.6	460	9	AA022681	AA022681 ze/1907.s
12	18	5.6	467	12	BG713416	BG713416 p911n.pk0
13	18	5.6	480	13	BM486606	BM486606 psm2n.pk0
14	18	5.6	485	12	BF850925	BF850925 MR4-EN007
15	18	5.6	512	12	BF550791	BF550791 UT-R-CO-1
16	18	5.6	535	17	BM435307	BM435307 BOHGU36TR
17	18	5.6	540	13	BM490160	BM490160 psp2n.pk0
18	18	5.6	550	13	BM491990	BM491990 psp2n.pk0
19	18	5.6	562	9	AJ443979	AJ443979 AJ443979
20	18	5.6	564	13	BM324433	BM324433 P1C1_31_B
21	18	5.6	569	13	BM487984	BM487984 psm2n.pk0
22	18	5.6	573	13	B1067932	B1067932 p911n.pk0
23	18	5.6	601	13	BM439457	BM439457 p911n.pk0
24	18	5.6	602	9	AL703781	AL703781 DRFP686T
25	18	5.6	604	13	BM489914	BM489914 psm2n.pk0
26	18	5.6	606	12	BG713718	BG713718 p911n.pk0
27	18	5.6	612	13	BM487383	BM487383 B1487383
28	18	5.6	621	13	BM489278	BM489278 psm2n.pk0
29	18	5.6	623	12	BG712411	BG712411 p911n.pk0
30	18	5.6	659	13	BM489885	BM489885 psm2n.pk0
31	18	5.6	663	13	B1395012	B1395012 psp1n.pk0
32	18	5.6	664	17	BM487055	BM487055 B0G1L27F
33	18	5.6	681	13	B1395000	B1395000 psp1n.pk0
34	18	5.6	689	9	AJ444876	AJ444876 AJ444876
35	18	5.6	696	17	AG064805	AG064805 Pan treg1
36	18	5.6	701	17	BH523619	BH523619 B0GK048TF
37	18	5.6	715	9	AJ397860	AJ397860 AJ397860
38	18	5.6	718	17	BH713237	BH713237 B0MNA43TF
39	18	5.6	733	9	AJ444936	AJ444936 AJ444936
40	18	5.6	758	17	BH504178	BH504178 B0E2C86TR
41	18	5.6	759	12	BG819881	BG819881 602781962
42	18	5.6	800	17	BM421309	BM421309 B0GPR22TF
43	18	5.6	841	13	B1562021	B1562021 603255073
44	18	5.6	846	9	AJ455107	AJ455107 AJ455107
45	18	5.6	870	14	BQ223367	BQ223367 AGENCOURT

## ALIGNMENTS

RESULT 1  
C91774 305 bp mRNA linear EST 04-APR-2002  
C91774 Rice panicle shorter than 3cm Oryza sativa (japonica  
LOCUS  
DEFINITION  
C91774.1 GI:3061142  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Oryza sativa (japonica cultivar-group).  
Oryza sativa (japonica cultivar-group).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriaraloideae; Oryzaceae; Oryza.  
1 (bases 1 to 305)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Sasaki, T. and Yamamoto, K.  
Rice CDNA from panicle  
Unpublished (1997)  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
PROJECT = 'RGP'  
POLY-A-NO.

FEATURES  
source  
1..305  
/organism="Oryza sativa (japonica cultivar-group)"

BASE COUNT 44 a 110 c 92 g 54 t 5 others

ORIGIN

Query Match 5.9%; Score 19; DB 14; Length 305;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GGAGCCAGCAGCCTGGCC 33  
192 GGAGCCAGCAGCCTGGCC 174

RESULT 2  
BI480603/c  
LOCUS  
DEFINITION BI480603 777 bp mRNA linear EST 28-AUG-2001  
EST0056 Habanero leaf cDNA Capsicum chinense clone Jn 51  
ACCESSION BI480603  
VERSION BI480603  
KEYWORDS EST.  
SOURCE Capsicum chinense.  
ORGANISM Capsicum chinense.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.  
1 (bases 1 to 777)  
Nevarez J.G. and O'Connell, M.A.  
Drought responsive transcripts in Capsicum chinense (habanero) leaves  
Unpublished (2001)  
Contact: O'Connell, MA  
Department of Agronomy and Horticulture  
New Mexico State University  
MSC 3Q, P.O. Box 30003, Las Cruces, NM 88003, USA  
Tel: 505 646 5172  
Fax: 505 646 6041  
Email: moconnell@nmsu.edu  
Insert Length: 1200 Std Error: 0.00  
Seq Primer: T3  
High quality sequence stop: 777.  
Location/Qualifiers  
1. 777  
/organism="Capsicum chinense"  
/cultivar="Early Scotch Bonnet"  
/db\_xref="taxon:80379"  
/clone="Jn 51"  
/clone\_lib="Habanero leaf cDNA"  
/dev\_stage="drought-stressed"  
/note="Organ: leaf; Vector: Uni-zap XR; Site\_1: EcoRI; Site\_2: Xho I."  
BASE COUNT 199 a 188 c 153 g 237 t  
ORIGIN

Query Match 5.9%; Score 19; DB 13; Length 777;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AGGCCAGCAGCCTGGCGA 35  
251 AGGCCAGCAGCCTGGCGA 233

RESULT 3  
BE791590  
LOCUS  
DEFINITION BE791590 1099 bp mRNA linear EST 20-SEP-2000  
601585915F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3940395 5',  
mRNA sequence.

ACCESSION BE791590  
VERSION BE791590.1 GI:10212788  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgs.nci.nih.gov/.  
1 (bases 1 to 1099)  
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
plate: LLCM790 row: m column: 04  
High quality sequence start: 7  
High quality sequence stop: 681.  
Location/Qualifiers  
1. 1099  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3940395"  
/clone\_lib="NIH\_MGC\_7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 221 a 321 c 342 g 215 t  
ORIGIN

Query Match 5.9%; Score 19; DB 12; Length 1099;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 GGGATCCGAGGACCCAC 183  
963 GGGATCCGAGGACCCAC 981

RESULT 4  
AI824830/c  
LOCUS  
DEFINITION AI824830 344 bp mRNA linear EST 16-DEC-1999  
WP02607.x1 NCI-CGAP-GC6 Homo sapiens cDNA clone IMAGE:2304493 3',  
similar to SN:HM62.CHICK P26584 HIGH MOBILITY GROUP PROTEIN HM62 ;,  
mRNA sequence.  
ACCESSION AI824830  
VERSION AI824830.1 GI:5445501  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.





QY 130 GGGATGTTCTGCGCCG 147  
 DB 299 GGGATGTTCTGCGCCG 316

RESULT 7  
 LOCUS BF086409 409 bp mRNA linear EST 19-OCT-2000  
 DEFINITION RC6-GN0070-120900-022-E10 GN0070 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF086409  
 VERSION BF086409.1 GI:10892119  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 409)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 JOURNAL MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=rc6-GN0070-120900-022-E10&f3=2000-09-12&f4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 13  
 High quality sequence stop: 387.  
 FEATURES  
 source  
 1. 409  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="GN0070"  
 /dev\_stage="Adult"  
 /note="Organ: placenta, normal; Vector: puc18; Site: 1: SmaI  
 products derived from ORESTES PCR (U.S. Letters Patent  
 application No. 196,716 - Ludwig Institute for Cancer  
 Research) profiles into the pUC 18 vector. Reverse  
 transcription of tissue mRNA and cDNA amplification were  
 performed under low stringency conditions."  
 BASE COUNT 90 a 125 c 109 g 85 t  
 ORIGIN

Query Match 5 6%; Score 18; DB 13; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 CCCACTGCGAGGAGCCCA 196  
 DB 293 CCCACTGCGAGGAGCCCA 310

RESULT 8  
 LOCUS BM487398/c 425 bp mRNA linear EST 07-FEB-2002  
 DEFINITION pgm2n.pk004.h1 Normalized Chicken Breast Muscle, Leg Muscle, and  
 Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA  
 clone pgm2n.pk004.h1 5' similar to gb|AAK15544.1 (AF346565) signal

ACCESSION BM487398  
 VERSION BM487398.1 GI:18608328  
 KEYWORDS EST.  
 SOURCE chicken.  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 425)  
 Cogburn, L.A. and Monsonego-Ornan, E.  
 ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and  
 Epiphyseal Growth Plate cDNA library, USDA/IRAFs Animal Genome  
 Project  
 Unpublished (2002)  
 JOURNAL MEDLINE 20202663  
 COMMENT Contact: Larry A. Cogburn  
 University of Delaware  
 Townsend Hall, Newark, DE 19717, USA  
 Tel: 302-831-1335  
 Fax: 302-831-2822  
 Email: cogburn@udel.edu, www.chickest.udel.edu.  
 FEATURES  
 source  
 1. 425  
 Location/Qualifiers  
 /organism="Gallus gallus"  
 /strain="Commercial broiler and Ottawa Res. Centre  
 Strains 90 & 21"  
 /db\_xref="taxon:9031"  
 /clone\_lib="pgm2n.pk004.h1"  
 /dev\_stage="E. coli EMDH10B"  
 /note="Vector: PCWVSPORT6; Library made from equivalent  
 pools of total RNA isolated from each tissue (embryonic  
 muscle 33.3%, juvenile muscle 33.3%, and epiphyseal growth  
 plate 33.3% of the final RNA pool). Single pass sequencing  
 from 5'-end of the final RNA pool."  
 /sex="Male and Female"  
 /tissue\_type="Breast muscle, leg muscle and epiphyseal  
 growth plate"  
 /dev\_stage="Breast, leg; Embryo (d19); post-hatch (d1, 1.3, 5, 7, 9,  
 11 weeks); growth plate (d1, d7, d14 post-hatch)"  
 /lab\_host="E. coli EMDH10B"  
 /note="Vector: PCWVSPORT6; Library made from equivalent  
 pools of total RNA isolated from each tissue (embryonic  
 muscle 33.3%, juvenile muscle 33.3%, and epiphyseal growth  
 plate 33.3% of the final RNA pool). Single pass sequencing  
 from 5'-end of the final RNA pool."  
 BASE COUNT 81 a 111 c 126 g 107 t  
 ORIGIN

Query Match 5.6%; Score 18; DB 13; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GGAGGCCAGCGCCTGGC 32  
 DB 86 GGAGGCCAGCGCCTGGC 69

RESULT 9  
 LOCUS BM491850/c 425 bp mRNA linear EST 07-FEB-2002  
 DEFINITION ppg2n.pk007.j9 Normalized Chicken Pituitary/Hypothalamus/Pineal  
 Library (ppg2n) Gallus gallus cDNA clone ppg2n.pk007.j9 5' similar  
 to gb|AAK15544.1 (AF346565) signal sequence receptor beta subunit  
 [Xenopus laevis], mRNA sequence.  
 ACCESSION BM491850  
 VERSION BM491850.1 GI:18612781  
 KEYWORDS EST.  
 SOURCE chicken.  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 425)  
 Porter, T.E. and Cogburn, L.A.  
 ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA  
 library, USDA/IRAFs Animal Genome Project  
 Unpublished (2002)

COMMENT Contact: Larry A. Cogburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES  
source  
1. .425  
/organism="Gallus gallus"  
/strain="Commercial broiler chickens"  
/db\_xref="taxon:9031"  
/clone="pdp2n.pk007.j9"  
/clone\_lib="Normalized Chicken  
Pituitary/Hypothalamus/Pineal Library (pdp2n)"  
/sex="Male and Female"  
/tissue\_type="Pituitary Gland/Hypothalamus/Pineal Gland"  
/dev\_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9  
weeks)"  
/lab\_host="E. coli EMDH10B"  
/note="Vector: PCWVSPORT6; Library made from equivalent  
pools of total RNA isolated from each tissue at different  
ages. Single pass sequencing from 5'-end"

BASE COUNT 79 a 112 c 125 g 104 t 5 others

ORIGIN  
Query Match 5.6%; Score 18; DB 13; Length 425;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GGAGGCCAGCAGCCTGCG 32  
|||||  
98 GGAGGCCAGCAGCCTGCG 81

Db

RESULT 10  
LOCUS B6713248/c 452 bp mRNA linear EST 08-MAY-2001  
DEFINITION pgl1n.pk001.d18 Normalized Liver Library Gallus gallus CDNA clone  
pgl1n.pk001.d18 5' similar to sp|P23438|SSRB\_CANRA  
TRANSLUCON-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)  
(SIGNAL SEQUENCE RECEPTOR BETA SUBUNIT) (SSR-BETA) (GP25H)  
p1r1a36679 signal sequence receptor beta chain precursor - dog,  
mRNA sequence.

ACCESSION B6713248  
VERSION B6713248.1 GI:14007198  
KEYWORDS chicken.  
SOURCE Gallus gallus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 452)  
Burnside,J., Morgan,R.W. and Cogburn,L.A.  
Chicken ESTs from a normalized liver library  
Unpublished (2001)  
Contact: Joan Burnside  
Molecular Endocrinology  
University of Delaware  
40 Townsend Hall, Newark, DE 19717, USA  
Tel: 302 831-1345  
Fax: 302-831-3411  
Email: joan@udel.edu, www.chickest.udel.edu.

FEATURES  
source  
1. .452  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/clone="pgl1n.pk001.d18"  
/clone\_lib="Normalized Liver Library"  
/sex="Male and Female"  
/tissue\_type="liver"  
/lab\_host="E.coli EMDH10B"  
/note="Vector: PCWVSPORT 6"

BASE COUNT 83 a 122 c 134 g 113 t

ORIGIN  
Query Match 5.6%; Score 18; DB 12; Length 452;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GGAGGCCAGCAGCCTGCG 32  
|||||  
113 GGAGGCCAGCAGCCTGCG 96

Db

RESULT 11  
LOCUS AA022681 460 bp mRNA linear EST 30-JAN-1997  
DEFINITION ze71907.s1 Soares.fetal\_heart\_NBHH19W Homo sapiens CDNA clone  
IMAGE:364476 3', mRNA sequence.

ACCESSION AA022681  
VERSION AA022681.1 GI:1486762  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 460)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B.,  
Chissoe,S., Dietrich,N., Dubuque,T., Favellio,A., Gish,W., Hawkins  
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,  
Schellenberg,K., Soares,M.B., Tan,F., Thierly-Meg,J., Treviskis,E.,  
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

TITLE JOURNAL  
MEDLINE  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LINT; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.  
Insert Length: 527 Std Error: 0.00  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 67.

FEATURES  
source  
1. .460  
/organism="Homo sapiens"  
/db\_xref="GDB:1281180"  
/db\_xref="taxon:9606"  
/clone="IMAGE:364476"  
/clone\_lib="Soares\_fetal\_heart\_NBHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT73D (pharmacia) with a  
modified polylinker; Site:1: Not 1; Site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTCACATCTGAGTGGAGCGCGCGCATCTTTTCTTTTCTT 3')  
TGTTCACATCTGAGTGGAGCGCGCGCATCTTTTCTTTTCTT 3')  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library was constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBHH19W."

BASE COUNT 152 a 75 c 98 g 132 t 3 others

ORIGIN  
Query Match 5.6%; Score 18; DB 9; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 AGAGGCCAGCAGCCTGG 31  
 Db 185 AGAGGCCAGCAGCCTGG 202

RESULT 12  
 LOCUS BG713416 467 bp mRNA linear EST 08-MAY-2001  
 DEFINITION pg11n.pk015.n15 Normalized Liver Library Gallus gallus cdna clone

TRANSLOCON-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)  
 (SIGMAL SEQUENCE RECEPTOR BETA SUBUNIT) (SSR-BETA) (GP25H)  
 p11a36679 signal sequence receptor beta chain precursor - dog,  
 mRNA sequence.

ACCESSION BG713416  
 VERSION BG713416.1 GI:14007366  
 KEYWORDS EST.  
 SOURCE chicken.  
 ORGANISM Gallus gallus

REFERENCE Archontology; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 PHASIANIDAE; Gallus.  
 1 (bases 1 to 467)  
 Burnside, J., Morgan, R.W. and Cogburn, L.A.  
 Chicken ESTs from a normalized liver library  
 Unpublished (2001)  
 CONTACT: Joan Burnside  
 Molecular Endocrinology  
 University of Delaware  
 40 Townsend Hall, Newark, DE 19717, USA  
 Tel: 302 831-1345  
 Fax: 302-831-3411  
 Email: joan@udel.edu, www.chickest.udel.edu.

FEATURES  
 source location/Qualifiers

1..467  
 /organism="Gallus gallus"  
 /db\_xref="taxon:9031"  
 /clone="pg11n.pk015.n15"  
 /clone\_lib="Normalized Liver Library"  
 /sex="Male and Female"  
 /tissue\_type="liver"  
 /lab\_host="E.coli EMDH108"  
 /note="Vector: PCWVSPORT 6"

BASE COUNT 84 a 128 c 135 g 117 t 3 others

Query Match 5.6%; Score 18; DB 12; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 GGAGGCCAGCAGCCTGGC 32  
 Db 129 GGAGGCCAGCAGCCTGGC 112

RESULT 13  
 LOCUS BM486606 490 bp mRNA linear EST 07-FEB-2002  
 DEFINITION pgm2n.pk002.c1 Normalized Chicken Breast Muscle, Leg Muscle, and

Epiphyseal Growth Plate CDNA Library (pgm2n) Gallus gallus cdna  
 clone pgm2n.pk002.c1 5' similar to gblAK1544.1 (AF346565) signal  
 sequence receptor beta subunit [Xenopus laevis], mRNA sequence.

ACCESSION BM486606  
 VERSION BM486606.1 GI:18607536  
 KEYWORDS EST.  
 SOURCE chicken.  
 ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 PHASIANIDAE; Gallus.  
 1 (bases 1 to 490)  
 Cogburn, L.A. and Monsonego-Orran, E.  
 ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and

JOURNAL  
 COMMENT Unpublished (2002)  
 Contact: Larry A. Cogburn  
 University of Delaware  
 Townsend Hall, Newark, DE 19717, USA  
 Tel: 302-831-1335  
 Fax: 302-831-2822  
 Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES  
 source location/Qualifiers

1..490  
 /organism="Gallus gallus"  
 /strain="Commercial broiler and Ottawa Res. Centre  
 Strains 90 & 21"  
 /db\_xref="taxon:9031"  
 /clone="pgm2n.pk002.c1"  
 /clone\_lib="Normalized Chicken Breast Muscle, Leg Muscle,  
 and Epiphyseal Growth Plate CDNA Library (pgm2n)"  
 /sex="Male and Female"  
 /tissue\_type="Breast muscle, leg muscle and epiphyseal  
 growth plate"  
 /dev\_stage="Breast, leg: Embryo(d19); post-hatch(1d,1.3,5,7,9  
 ,11 weeks); growth plate(1d,7d,14d post-hatch)"  
 /lab\_host="E. coli EMDH108"  
 /note="Vector: PCWVSPORT6; Library made from equivalent  
 pools of total RNA isolated from each tissue (embryonic  
 muscle 33.3%, juvenile muscle 33.3%, and epiphyseal growth  
 plate 33.3% of the final RNA pool). Single pass sequencing  
 from 5'-end"

BASE COUNT 88 a 132 c 147 g 123 t

Query Match 5.6%; Score 18; DB 13; Length 490;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 GGAGGCCAGCAGCCTGGC 32  
 Db 98 GGAGGCCAGCAGCCTGGC 81

RESULT 14  
 LOCUS BF850925 495 bp mRNA linear EST 16-JAN-2001  
 DEFINITION MR4-EN0075-241100-001-f07 EN0075 Homo sapiens cdna, mRNA sequence.  
 ACCESSION BF850925  
 VERSION BF850925.1 GI:12238087  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 495)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&ct2=MR4-EN0075-241100-001-f07&ct3=2000-11-24&ct4=1)

Seq primer: puc 18 forward  
High quality sequence start: 24  
High quality sequence stop: 495.  
Location/Qualifiers

## FEATURES

source

1. 495  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="EN0075"  
/dev\_stage="Adult"

/note="Organ: lung\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORSTRES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## BASE COUNT

112 a 126 c 152 g 105 t

Query Match  
Best Local Similarity 100.0%; Score 18; DB 12; Length 495;  
Pred. No. 1.6e+02; Mismatches 0; Gaps 0;

179 CCCACTGCGAGAGGCCA 196

129 CCCACTGCGAGAGGCCA 112

## RESULT 15

BF550791

LOCUS

DEFINITION

BF550791

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 512)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

CONTACT: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

CDNA Library Preparation: M.B. Soares Lab Clone Distribution:

Clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LNL (info@image.llnl.gov). IMAGE ID= 1774925

Seq primer: M13 forward.

Location/Qualifiers

1. 512

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-C0-jp-e-05-0-UI"

/clone\_id="UI-R-C0"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-C0 library is a subtracted library derived from the UI-R-A1

BASE COUNT 146 a 142 c 134 g 88 t 2 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 18; DB 12; Length 512;

Pred. No. 1.6e+02; Mismatches 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0;

13 GAGGAGCCAGCAGCCTG 30

378 GAGGAGCCAGCAGCAGCCTG 395

Search completed: June 2, 2003, 09:04:39  
Job time: 1516 secs

and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996).





SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 5.3%; Score 17; DB 4; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 GCCGACCTTGCGCCGA 82  
DB 572583 GCCGACCTTGCGCCGA 572599

RESULT 3  
US-08-675-508-13/C  
Sequence 13; Application US/08675508

PATENT No. 5856136  
GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
TITLE OF INVENTION: NOVEL HUMAN STEM CELL ANTIGENS

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive  
CITY: Palo Alto

STATE: CA  
COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/675,508

FILING DATE: Filed Herewith

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0066 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

IMMEDIATE SOURCE:  
LIBRARY: BRSTN0703

CLONE: 641178

US-08-675-508-13

Query Match 5.0%; Score 16; DB 2; Length 261;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 CCAGCAGCCTGGCCGA 35  
DB 28 CCAGCAGCCTGGCCGA 13

RESULT 4  
US-08-675-508-10/C  
Sequence 10; Application US/08675508

PATENT No. 5856136  
GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
TITLE OF INVENTION: NOVEL HUMAN STEM CELL ANTIGENS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive  
CITY: Palo Alto

STATE: CA  
COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/675,508

FILING DATE: Filed Herewith

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0066 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 262 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

IMMEDIATE SOURCE:  
LIBRARY: BRSTTUT01

CLONE: 604702

US-08-675-508-10

Query Match 5.0%; Score 16; DB 2; Length 262;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 CCAGCAGCCTGGCCGA 35  
DB 28 CCAGCAGCCTGGCCGA 13

RESULT 5

US-08-675-508-16/C

Sequence 16; Application US/08675508

PATENT No. 5856136  
GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
TITLE OF INVENTION: NOVEL HUMAN STEM CELL ANTIGENS

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive  
CITY: Palo Alto

STATE: CA  
COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/675,508

FILING DATE: Filed Herewith

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0066 US



```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 266 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: LUNGN0T03
CLONE: 728784
US-08-675-508-16

Query Match      5.0%; Score 16; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      20 CCAGCAGCCTGGCCGA 35
DB      42 CCAGCAGCCTGGCCGA 27

RESULT 6
US-08-675-508-9/c
; Sequence 9, Application US/08675508
; Patent No. 5856136
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: NOVEL HUMAN STEM CELL ANTIGENS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/675,508
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0066 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: HMT2RAT01
CLONE: 486681
US-08-675-508-9

Query Match      5.0%; Score 16; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      20 CCAGCAGCCTGGCCGA 35
DB      27 CCAGCAGCCTGGCCGA 12
```

```
RESULT 7
US-08-675-508-11/c
; Sequence 11, Application US/08675508
; Patent No. 5856136
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: NOVEL HUMAN STEM CELL ANTIGENS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
US-08-675-508-11

Query Match      5.0%; Score 16; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      20 CCAGCAGCCTGGCCGA 35
DB      27 CCAGCAGCCTGGCCGA 12

RESULT 8
US-08-675-508-12/c
; Sequence 12, Application US/08675508
; Patent No. 5856136
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: NOVEL HUMAN STEM CELL ANTIGENS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
US-08-675-508-12
```

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/675,508  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0066 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 335 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
IMMEDIATE SOURCE:  
LIBRARY: BRSTNOT03  
CLONE: 637479  
US-08-675-508-12

Query Match 5.0%; Score 16; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CCAGCAGCCTGGCCGA 35  
DB 27 CCAGCAGCCTGGCCGA 12

RESULT 9  
US-08-746-397-1/c  
; Sequence 1, Application US/08746397  
; Patent No. 6130061  
; GENERAL INFORMATION:  
; APPLICANT: NI, ET AL.  
; TITLE OF INVENTION: Human Stem Cell Antigen 2  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLIAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/746,397  
; FILING DATE: 11/8/96  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/007,287  
; FILING DATE: 11/9/95  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-494  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1163 BASE PAIRS  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: CDNA

US-08-746-397-1

Query Match 5.0%; Score 16; DB 3; Length 1163;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 CCAGCAGCCTGGCCGA 35  
DB 70 CCAGCAGCCTGGCCGA 55

RESULT 10  
US-08-465-380-115  
; Sequence 115, Application US/08465380  
; Patent No. 5863894  
; GENERAL INFORMATION:  
; APPLICANT: George P. Vlausk, Patric H. Stanssens,  
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
; APPLICANT: Yannick G.J. Ganssemans, Matthew Moyle,  
; APPLICANT: Peter W. Bergum  
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
; NUMBER OF SEQUENCES: 356  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,380  
FILING DATE: June 5, 1995  
CLASSIFICATION: 350  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-465-380-115

Query Match 4.7%; Score 15; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GATGCCGACTTGGC 77  
DB 30 GATGCCGACTTGGC 44

RESULT 11  
US-08-480-478-30  
; Sequence 30, Application US/08480478

Patent No. 5864009  
GENERAL INFORMATION:  
APPLICANT: GEORGE P. VLASUK; PATRICK ERIC  
APPLICANT: HUGO STANSSENS; JORIS HILDA  
APPLICANT: LIEVEN MESSENS; MARC JOZEF  
APPLICANT: LAUMERYS; YVES RENE LAROCHE;  
APPLICANT: LAURENT STEPHANE JESPEERS; and  
APPLICANT: YANNICK GEORGES JOZEF  
APPLICANT: GANSEMAN  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTI-  
TITLE OF INVENTION: COAGULANT PROTEIN  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,478  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: 18 OCTOBER 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 208/290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-480-478-30  
Query Match 4.7%; Score 15; DB 2: Length 45;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 63 GATGCCGCGACTGGC 77  
|||||  
DB 30 GATGCCGCGACTGGC 44

RESULT 12  
US-08-486-397-115  
Sequence 115, Application US/08486397  
Patent No. 5865542  
GENERAL INFORMATION:  
APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Menssens, Marc J. Lauwereys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 357  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,397  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/269  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-486-397-115  
Query Match 4.7%; Score 15; DB 2: Length 45;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 63 GATGCCGCGACTGGC 77  
|||||  
DB 30 GATGCCGCGACTGGC 44

RESULT 13  
US-08-486-399-115  
Sequence 115, Application US/08486399  
Patent No. 5865543  
GENERAL INFORMATION:  
APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Menssens, Marc J. Lauwereys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,399

FILED DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 213/270  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-486-399-115

Query Match  
Best Local Similarity 100.0%; Score 15; DB 2; Length 45;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GATGCCGACTTGGC 77  
|||||  
DB 30 GATGCCGACTTGGC 44

RESULT 14  
US-08-461-965-115  
Sequence 115, Application US/08461965  
Patent No. 5872098  
GENERAL INFORMATION:  
APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
APPLICANT: Joris H.L. Mensens, Marc J. Lauwerys,  
APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
APPLICANT: Yannick G.J. Ganssemans, Matthew Moyle,  
APPLICANT: Peter W. Bergum  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
NUMBER OF SEQUENCES: 356  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,965  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/326,110  
FILING DATE: October 18, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 210/243  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 115:

SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-461-965-115

Query Match  
Best Local Similarity 100.0%; Score 15; DB 2; Length 45;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GATGCCGACTTGGC 77  
|||||  
DB 30 GATGCCGACTTGGC 44

RESULT 15  
US-08-326-110A-30  
Sequence 30, Application US/08326110A  
Patent No. 5945275  
GENERAL INFORMATION:

APPLICANT: GEORGE P. VLASUK; PATRICK ERIC  
APPLICANT: HUGO STANSSENS; JORIS HILDA  
APPLICANT: LIEVEN MESSENS; MARC JOZER  
APPLICANT: LAUWERYS; YVES RENE LAROCHE;  
APPLICANT: LAURENT STEPHANE JESPEERS; and  
APPLICANT: YANNICK GEORGES JOZER  
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTI-  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/326,110A  
FILING DATE: 18 OCTOBER 1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: BIGGS, SUZANNE L.  
REGISTRATION NUMBER: 30,158  
REFERENCE/DOCKET NUMBER: 208/290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-326-110A-30

Query Match  
Best Local Similarity 100.0%; Score 15; DB 2; Length 45;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GATGCCGACTTGGC 77

Mon Jun 2 10:57:27 2003

us-09-977-137a-3.oli.rni

Page 7

Db 30 GATGGCCGACTTGGC 44

Search completed: June 2, 2003, 09:05:42  
Job time : 52 secs



Mon Jun 2 10:57:27 2003

us-09-977-137a-3.oli.rnpb

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 08:35:40 ; Search time 136 Seconds  
(without alignments)  
3182.538 Million cell updates/sec

Title: US-09-977-137A-3  
Perfect score: 321  
Sequence: 1 atgacacactgcgagagagc.....agggaatgtctcctgcgcg 321

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 845702 seqs, 674182571 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEM\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCRTS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	5.3	463	9	US-09-918-995-19156
2	17	5.3	484	9	US-09-918-995-20836
3	17	5.3	1356	9	US-09-712-363-25
4	16	5.0	151	9	US-09-796-692-5769
5	16	5.0	151	9	US-10-040-862-5769
6	16	5.0	154	9	US-09-796-692-5723
7	16	5.0	154	9	US-10-040-862-5723
8	16	5.0	158	9	US-09-796-692-5737
9	16	5.0	158	9	US-09-796-692-5737
10	16	5.0	158	9	US-09-796-692-5737
11	16	5.0	158	9	US-10-040-862-5667
12	16	5.0	158	9	US-10-040-862-5667
13	16	5.0	158	9	US-10-040-862-5667
14	16	5.0	159	9	US-09-796-692-6280
15	16	5.0	159	9	US-10-020-758-25
16	16	5.0	261	10	US-09-934-586A-13
17	16	5.0	262	10	US-09-934-586A-10
18	16	5.0	262	10	US-09-934-586A-16
19	16	5.0	266	10	US-09-934-586A-16

20	16	5.0	278	10	US-09-934-586A-9	Sequence 9, Appli
c 21	16	5.0	289	10	US-09-934-586A-11	Sequence 11, Appli
c 22	16	5.0	335	10	US-09-934-586A-12	Sequence 12, Appli
c 23	16	5.0	365	10	US-09-867-550-653	Sequence 653, App
c 24	16	5.0	395	10	US-09-983-965-3887	Sequence 3887, App
c 25	16	5.0	405	10	US-09-960-352-2232	Sequence 2232, App
c 26	16	5.0	474	9	US-09-918-995-17595	Sequence 17595, A
c 27	16	5.0	495	9	US-09-918-995-13450	Sequence 13450, A
c 28	16	5.0	1163	12	US-10-101-747-1	Sequence 1, Appli
c 29	16	5.0	1323	9	US-09-738-626-1656	Sequence 1656, App
c 30	16	5.0	1739	9	US-09-804-073-1	Sequence 105, App
c 31	16	5.0	6232	9	US-10-098-841-105	Sequence 1590, App
c 32	16	5.0	17761	9	US-10-092-154-1596	Sequence 1596, App
c 33	16	5.0	17761	9	US-09-764-847-1596	Sequence 3, Appli
c 34	16	5.0	34668	9	US-09-900-449A-3	Sequence 1, Appli
c 35	16	5.0	3509400	9	US-09-738-626-1	Sequence 2556, App
c 36	15	4.7	138	10	US-09-923-876-3471	Sequence 474, App
c 37	15	4.7	226	10	US-09-923-876-2556	Sequence 4927, App
c 38	15	4.7	272	9	US-09-925-299-474	Sequence 1856, App
c 39	15	4.7	272	10	US-09-925-299-474	Sequence 2269, App
c 40	15	4.7	280	10	US-09-923-876-4927	Sequence 748, App
c 41	15	4.7	302	10	US-09-833-381-1856	Sequence 6603, App
c 42	15	4.7	383	10	US-09-783-590-2269	Sequence 10733, A
c 43	15	4.7	397	10	US-09-864-761-748	
c 44	15	4.7	423	10	US-09-783-590-6503	
c 45	15	4.7	444	10	US-09-960-352-10733	

# ALIGNMENTS

RESULT 1  
US-09-918-995-19156  
; Sequence 19156, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918, 995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235, 076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19156  
; LENGTH: 463  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: (1)...(463)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-19156

Query Match 5.3%; Score 17; DB 9; Length 463;  
Best Local Similarity 100.0%; Pred. No. 7.5;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 CCGGCTGTCTGAAC TC 105  
|||||  
Db 58 CCGGCTGTCTGAAC TC 74

RESULT 2  
US-09-918-995-20836  
; Sequence 20836, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES

```

FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 20836
LENGTH: 484
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(484)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-20836

```

```

Query Match
Best Local Similarity 5.3%; Score 17; DB 9; Length 484;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 89 CCCTGCTGTGACTC 105
64 CCCTGCTGTGACTC 80

```

```

RESULT 3
US-09-712-363-25
Sequence 25, Application US/09/12363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
PRIOR FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 1356
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-25

```

```

Query Match
Best Local Similarity 5.3%; Score 17; DB 9; Length 1356;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 66 GCGCGACTTGCGCGCA 82
876 GCGCGACTTGCGCGCA 892

```

```

RESULT 4
US-09-796-692-5769/C
Sequence 5769, Application US/09/796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5769
LENGTH: 151
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (43)
OTHER INFORMATION: (43)
US-09-796-692-5769

```

```

Query Match
Best Local Similarity 5.0%; Score 16; DB 9; Length 151;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 166 GGATCCTCAGGACCC 181
22 GGATCCTCAGGACCC 7

```

```

RESULT 5
US-10-040-862-5769/C
Sequence 5769, Application US/10040862
Publication No. US20030078396A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
FILE REFERENCE: 014058-01352005
CURRENT APPLICATION NUMBER: US/10/040,862

```



Mon Jun 2 10:57:27 2003

us-09-977-137a-3.oli.rnpb

Page 3

```

; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5769
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (43)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-5769

Query Match          5.0%; Score 16; DB 9; Length 151;
Best Local Similarity 100.0%; Pred No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      166 GGATCCTCAGGACCC 181
DB      22 GGATCCTCAGGACCC 7

RESULT 6
US-09-796-692-5723/C
; Sequence 5723, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5723
```

```

; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5723
; LENGTH: 154
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-5723

Query Match          5.0%; Score 16; DB 9; Length 154;
Best Local Similarity 100.0%; Pred No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      166 GGATCCTCAGGACCC 181
DB      22 GGATCCTCAGGACCC 7

RESULT 7
US-10-040-862-5723/C
; Sequence 5723, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5723
```

LENGTH: 154  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-040-862-5723

Query Match  
Best Local Similarity 5.0%; Score 16; DB 9; Length 154;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGATCCTCAGGACCC 181  
DB 22 GGATCCTCAGGACCC 7

RESULT 8  
US-09-796-692-3737  
Sequence 3737, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander  
APPLICANT: Mannion, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
PRIOR FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3737  
LENGTH: 158  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-3737

Query Match  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGATCCTCAGGACCC 181  
DB 133 GGATCCTCAGGACCC 148

RESULT 9  
US-09-796-692-5607  
Sequence 5607, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:

APPLICANT: Galger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
PRIOR FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5607  
LENGTH: 158  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-5607

Query Match  
Best Local Similarity 5.0%; Score 16; DB 9; Length 158;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGATCCTCAGGACCC 181  
DB 133 GGATCCTCAGGACCC 148

RESULT 10  
US-09-796-692-5667  
Sequence 5667, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander  
APPLICANT: Mannion, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
PRIOR FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999

```
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; NUMBER OF SEQ ID NOS: 9597
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 5667
;; LENGTH: 158
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-976-692-5667
```

```
Query Match      5.0%; Score 16; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      166 GGATCCTCAGGCACC 181
        |||
Db       133 GGATCCTCAGGCACC 148
```

## RESULT 11

```
US-10-040-862-3737
;; Sequence 3737, Application US/10040862
;; Publication No. US20030078396A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Retter, Marc
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;; FILE REFERENCE: 014058-013520US
;; CURRENT FILING DATE: 2001-11-06
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: US 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: US 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; PRIOR APPLICATION NUMBER: US 09/796,692
;; PRIOR FILING DATE: 2001-03-01
;; NUMBER OF SEQ ID NOS: 10467
;; SOFTWARE: FastSeq for Windows Version 3.0
```

```
;; SEQ ID NO 3737
;; LENGTH: 158
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-040-862-3737
```

```
Query Match      5.0%; Score 16; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      166 GGATCCTCAGGCACC 181
        |||
Db       133 GGATCCTCAGGCACC 148
```

## RESULT 12

```
US-10-040-862-5607
;; Sequence 5607, Application US/10040862
;; Publication No. US20030078396A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Retter, Marc
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
;; FILE REFERENCE: 014058-013520US
;; CURRENT FILING DATE: 2001-11-06
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: US 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: US 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; PRIOR APPLICATION NUMBER: US 09/796,692
;; PRIOR FILING DATE: 2001-03-01
;; NUMBER OF SEQ ID NOS: 10467
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 5607
;; LENGTH: 158
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-040-862-5607
```

```
Query Match      5.0%; Score 16; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      166 GGATCCTCAGGCACC 181
        |||
Db       133 GGATCCTCAGGCACC 148
```

RESULT 13

```

US-10-040-862-5667
; Sequence 5667, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-01352005
; CURRENT APPLICATION NUMBER: US/10/040, 862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5667
; LENGTH: 158
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-5667

```

Query Match

Best Local Similarity 100.0%; Score 16; DB 9; Length 158;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGATCCTCAGCAGCC 181  
DB 133 GGATCCTCAGCAGCC 148

RESULT 14

```

US-09-796-692-6280
; Sequence 6280, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01

```

```

; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6280
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n-A,T,C or G
US-09-796-692-6280

```

Query Match

Best Local Similarity 100.0%; Score 16; DB 9; Length 159;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 GGATCCTCAGCAGCC 181  
DB 134 GGATCCTCAGCAGCC 149

RESULT 15

```

US-10-040-862-6280
; Sequence 6280, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
; FILE REFERENCE: 014058-01352005
; CURRENT APPLICATION NUMBER: US/10/040, 862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201

```

```

; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6280
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-6280

```

```

Query Match          5.0%; Score 16; DB 9; Length 159;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      166 GGATCTCTCAGGCACCC 181
          ||||||||||||||||
DB      134 GGATCTCTCAGGCACCC 149

```

Search completed: June 2, 2003, 09:48:27  
 Job time : 137 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:29:22 ; Search time 74 Seconds  
(without alignments)  
210.680 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 117  
Sequence: 1 MHCEFASSLAEHKLDVRE.....HARKGNVSCPSAMSHPOFEK 117

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	100.0	117	23	AAU97552
2	64	54.7	117	23	AAU97560
3	57	48.7	117	23	AAU97553
4	57	48.7	117	23	AAU97555
5	57	48.7	117	23	AAU97556
6	57	48.7	117	23	AAU97557
7	57	48.7	117	23	AAU97558
8	57	48.7	117	23	AAU97559
9	55	47.0	144	23	AAU97551
10	48	41.0	118	23	AAU97554

11	29	24.8	144	20	AAU01816
12	24	20.5	159	15	AAU49668
13	10	8.5	10	22	AAU97662
14	10	8.5	10	23	AAU80475
15	8	6.8	8	19	AAU59212
16	8	6.8	8	22	AAU35433
17	8	6.8	8	22	AAU35442
18	8	6.8	8	22	AAU68616
19	8	6.8	8	23	AAU77486
20	8	6.8	8	23	AAU57464
21	8	6.8	10	22	AAU99027
22	8	6.8	205	20	AAU93967
23	8	6.8	254	20	AAU93969
24	8	6.8	254	22	AAU46421
25	8	6.8	396	20	AAU93966
26	8	6.8	396	22	AAU46420
27	8	6.8	396	22	AAU46423
28	8	6.8	448	22	AAU46425
29	8	6.8	659	22	AAU46426
30	8	6.8	659	22	AAU46427
31	7	6.0	158	21	AAU15367
32	7	6.0	161	22	AAU21632
33	7	6.0	165	21	AAU15366
34	7	6.0	179	21	AAU15365
35	7	6.0	192	22	ABU10936
36	7	6.0	192	22	AAU96372
37	7	6.0	192	22	AAU21776
38	7	6.0	359	21	AAU42524
39	7	6.0	378	21	AAU42523
40	7	6.0	515	21	AAU42522
41	7	6.0	1122	22	ABU61604
42	7	6.0	2659	22	ABU68426
43	6	5.1	19	21	AAU91236
44	6	5.1	19	23	ABU684205
45	6	5.1	25	23	AAU88121

#### ALIGNMENTS

RESULT 1  
AAU97552  
AAU97552 standard; Protein: 117 AA.

XX  
AC AAU97552;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Synthetic Merr chelon variant.  
XX  
KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW Irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein; Merr.  
XX  
OS Synthetic.  
XX  
PN WO200230962-A2.  
XX  
PD 18-APR-2002.  
XX  
PF 12-OCT-2001; 2001WO-US31819.  
XX  
PR 12-OCT-2000; 2000US-240465P.  
XX  
PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Summers AO, Caguiat JJ;  
XX  
DR WPI, 2002-435437/46.  
XX  
DR N-PSDB; ABK52207, ABK52213.  
XX

Merr protein which  
protein product of  
influenza haemaggl  
peptide STREP tag.  
Streptavidin tagge  
Epitope peptide #3  
Nascent protein de  
STREP-tag II seque  
AAV-helper Plasmid  
STREP tag II pept  
Streptavidin detec  
Plasmid pBBP21 pro  
Plasmid pBBP22 pro  
Bilin binding-prot  
Plasmid pBBP20 pro  
Bilin binding-prot  
Bilin binding-prot  
Bilin binding-prot  
Bilin binding-prot  
Bilin binding-prot  
Arabidopsis thalia  
Novel human neopla  
Arabidopsis thalia  
Arabidopsis thalia  
Human ovarian and/  
Human reproductive  
Novel human neopla  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Drosophila melanog  
Modified measles v  
Measles virus idea  
Insulin/Insulin-II

PT Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
XX  
PS Claim 2; Page 20; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydropionic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents the synthetic MerR chelon variant protein of the  
CC invention. This sequence is one of the heavy metal binding proteins  
CC termed chelons of the invention.  
CC Note: ABRK2207 encodes only amino acids 1-107 of this protein.

XX Sequence 117 AA:

Query Match 100.0%; Score 117; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 3.1e-115;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTHCEASSLAHEHKLKDYREKMDLARMETVLSLVACAHARGNVSCPTLASLOGSSGT 60  
DB 1 MTHCEASSLAHEHKLKDYREKMDLARMETVLSLVACAHARGNVSCPTLASLOGSSGT 60

OY 61 HCEASSLAHEHKLKDYREKMDLARMETVLSLVACAHARGNVSCPSAMSHPOPEK 117  
DB 61 HCEASSLAHEHKLKDYREKMDLARMETVLSLVACAHARGNVSCPSAMSHPOPEK 117

RESULT 2

AAU97560 standard; Protein; 117 AA.

AAU97560;

13-AUG-2002 (first entry)

Synthetic cadmium/mercury ion binding chelon protein #8.

KW Mercuric ion; contaminated soil; ground water; hydropionic solution;  
KW Irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.

Synthetic.

WO200230962-A2.

18-APR-2002.

12-OCT-2001; 2001WO-US31819.

12-OCT-2000; 2000US-240465P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
PA  
XX  
PI Summers AO, Caguiat JI;  
XX  
DR WPI; 2002-435437/46.

PT Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
XX  
PS Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydropionic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AAU97553-AAU97560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.

XX Sequence 117 AA:

Query Match 54.7%; Score 64; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 1.8e-59;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 LOGSSGTHCEASSLAHEHKLKDYREKMDLARMETVLSLVACAHARGNVSCPSAMSHP 113

DB 54 LOGSSGTHCEASSLAHEHKLKDYREKMDLARMETVLSLVACAHARGNVSCPSAMSHP 113

OY 114 QPEK 117

DB 114 QPEK 117

RESULT 3

AAU97553 standard; Protein; 117 AA.

AAU97553;

13-AUG-2002 (first entry)

Synthetic cadmium/mercury ion binding chelon protein #1.

KW Mercuric ion; contaminated soil; ground water; hydropionic solution;  
KW Irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.

Synthetic.



PN WO200230962-A2.  
XX 18-APR-2002.  
XX 12-OCT-2001; 2001WO-US31819.  
XX 12-OCT-2000; 2000US-240465P.  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX Sumners AO, Caguiat JJ;  
XX WPI; 2002-435437/46.  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
XX Claim 4; Page 22; 42pp; English.  
XX The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AAU97553-AAU97560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.  
XX  
XX Sequence 117 AA;  
SQ  
Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 4.2e-52;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 SSIAEHKLVKDYREKMDLAREMTVLSLVCACHARKGNVSCPLIASLGSGTHCEE 64  
DB 8 SSIAEHKLVKDYREKMDLAREMTVLSLVCACHARKGNVSCPLIASLGSGTHCEE 64  
RESULT 4  
AAU97555 standard; Protein: 117 AA.  
XX  
XX AAU97555;  
AC  
DT 13-AUG-2002 (first entry)  
XX  
XX Synthetic cadmium/mercury ion binding chelon protein #3.  
DE  
XX Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW irrigating water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.

XX  
OS Synthetic.  
XX WO200230962-A2.  
XX 18-APR-2002.  
XX 12-OCT-2001; 2001WO-US31819.  
XX 12-OCT-2000; 2000US-240465P.  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX Sumners AO, Caguiat JJ;  
XX WPI; 2002-435437/46.  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
XX Claim 4; Page 22; 42pp; English.  
XX The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AAU97553-AAU97560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.  
XX  
XX Sequence 117 AA;  
SQ  
Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 4.2e-52;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 MADLAREMTVLSLVCACHARKGNVSCPLIASLGSGTHCEEASIAEHKLVKDYRE 78  
DB 22 MADLAREMTVLSLVCACHARKGNVSCPLIASLGSGTHCEEASIAEHKLVKDYRE 78  
RESULT 5  
AAU97556 standard; Protein: 117 AA.  
XX  
XX AAU97556;  
AC  
DT 13-AUG-2002 (first entry)  
XX  
XX Synthetic cadmium/mercury ion binding chelon protein #4.  
DE  
XX Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW irrigating water; waste stream; contaminated aqueous medium;

AAU97558;

DT 13-AUG-2002 (first entry)  
XX Synthetic cadmium/mercury ion binding chelon protein #6.  
DE  
XX  
XX  
KM Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KM Irrigation water; waste stream; contaminated aqueous medium;  
KM biological fluid; gastrointestinal tract; chelon protein;  
KM enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.  
XX Synthetic.  
XX  
XX WO200230962-A2.  
XX  
XX  
XX 18-APR-2002.  
XX  
XX 12-OCT-2001; 2001WO-US31819.  
XX  
XX 12-OCT-2000; 2000US-240465P.  
XX  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
XX Summers AO, Caguiat JT;  
XX  
XX WPI; 2002-435437/46.  
XX  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
XX  
PS Claim 4; Page 22; 42pp; English.  
XX  
XX The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AA097553-AA097560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.  
XX  
XX Sequence 117 AA:  
SQ  
Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 4.2e-52;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 ETVLSTLVACARARKGNVSCPLIASLOGSSGTHCEFASSLAHEHLKDYREKMAADLAR 85  
Db 29 ETVLSTLVACARARKGNVSCPLIASLOGSSGTHCEFASSLAHEHLKDYREKMAADLAR 85  
RESULT 8  
AA097559  
ID AA097559 standard; Protein; 117 AA.

XX  
AC AA097559;  
DT 13-AUG-2002 (first entry)  
XX Synthetic cadmium/mercury ion binding chelon protein #7.  
DE  
XX  
XX  
KM Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KM Irrigation water; waste stream; contaminated aqueous medium;  
KM biological fluid; gastrointestinal tract; chelon protein;  
KM enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.  
XX Synthetic.  
XX  
XX WO200230962-A2.  
XX  
XX  
XX 18-APR-2002.  
XX  
XX 12-OCT-2001; 2001WO-US31819.  
XX  
XX 12-OCT-2000; 2000US-240465P.  
XX  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
XX Summers AO, Caguiat JT;  
XX  
XX WPI; 2002-435437/46.  
XX  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
XX  
PS Claim 4; Page 22; 42pp; English.  
XX  
XX The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AA097553-AA097560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.  
XX  
XX Sequence 117 AA:  
SQ  
Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 4.2e-52;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 CPLIASLOGSSGTHCEFASSLAHEHLKDYREKMAADLARMTYVSELYCACHARKGNV 104  
Db 48 CPLIASLOGSSGTHCEFASSLAHEHLKDYREKMAADLARMTYVSELYCACHARKGNV 104

ID	AAU97551	standard; Protein; 144 AA.
XX	AAU97551	
AC	AAU97551;	
DT	13-AUG-2002	(first entry)
XX		
DE	Shigella flexneri wild-type Merr protein.	
KM	Mercuric ion; contaminated soil; ground water; hydroponic solution;	
KW	Irrigation water; waste stream; contaminated aqueous medium;	
KW	biological fluid; gastrointestinal tract; chelon protein;	
KW	enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;	
KW	heavy metal binding protein; MeRR.	
OS	Shigella flexneri.	
PN	WO200230962-A2.	
PD	18-APR-2002.	
PF	12-OCT-2001; 2001MO-US31819.	
PR	12-OCT-2000; 2000US-240465P.	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.	
PI	Summers AO, Caguiat JJ;	
DR	MPi: 2002-435437/46.	
N-PSDB; ABR52206.		
PT	Novel non-naturally occurring recombinant DNA molecule encoding a	
PT	chelon protein useful for binding divalent cation mercury from	
PT	contaminated soil, water, aqueous medium including biological fluids -	
PS	Disclosure; Page 20; 42pp; English.	
CC	The present invention relates to a new non-naturally occurring	
CC	recombinant DNA molecule comprising a sequence encoding a chelon protein	
CC	which binds mercuric ions. The invention is useful for recombinantly	
CC	producing a protein in a host-cell, by infecting or transforming a host	
CC	cell capable of expressing a chelon coding sequence with a vector	
CC	comprising a promoter active in the host cell operably linked to a coding	
CC	region for the protein to produce a recombinant host cell and culturing	
CC	the recombinant host cell under conditions, where DNA is expressed.	
CC	The nucleic acid encoding the chelon protein is useful for binding	
CC	divalent mercuric ions, to take up, sequester and concentrate the heavy	
CC	metal ions from contaminated soil, ground water, hydroponic solutions or	
CC	irrigation water of waste streams. The DNA of the invention, when	
CC	immobilised onto a solid support, is useful for concentrating heavy metal	
CC	ions from contaminated environment waste streams or contaminating	
CC	aqueous medium including biological fluids. The nucleic acid, when	
CC	recombinantly expressed in enteric bacteria (which are nontoxicogenic and	
CC	nonpathogenic), is suitable for use in the in vivo sequestration and	
CC	elimination of mercuric ion from gastrointestinal tracts of animals or	
CC	humans exposed to toxic metal ions such as mercury and/or cadmium. The	
CC	molecules of the invention are also useful in water treatment resins.	
CC	The nucleic acid of the invention is highly specific and binds divalent	
CC	cation such as mercury or cadmium with high affinity. The present amino	
CC	acid sequence represents the Shigella flexneri wild-type Merr protein of	
CC	the invention. This sequence was used in the methods of the invention for	
CC	production of heavy metal binding proteins termed chelons.	
SQ	Sequence 144 AA;	
Query Match	47.0%; Score 55; DB 23; Length 144;	
Best Local Similarity	100.0%; Pred. No. 6,4e-50;	
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		

Db 80 THEEFASSLAHEHTKIDYREKADLAREMYLSELVCACHARKGNVSOPLIASLQG 134

RESULT 10  
AA097554  
ID AA097554 standard; Protein; 118 AA.  
XX  
XX AA097554;  
AC  
XX  
XX 13-AUG-2002 (first entry)  
DT  
XX  
XX  
DE Synthetic cadmium/mercury ion binding chelon protein #2.  
XX  
XX  
KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW Irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.  
XX  
XX  
OS Synthetic.  
OS  
PN WO200230962-A2.  
PN  
PD 18-APR-2002.  
PD  
XX 12-OCT-2001; 2001WO-US31819.  
XX PF  
XX 12-OCT-2000; 2000US-240465P.  
PR  
XX 12-OCT-2000; 2000US-240465P.  
PA  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
PA  
PI Summers AO, Cagulat JI;  
PI  
XX WPI: 2002-435437/46.  
XX  
XX  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
PT  
PS Claim 4; Page 22; 42pp; English.

The present invention relates to a new non-naturally occurring recombinant DNA molecule comprising a sequence encoding a chelon protein which binds mercuric ions. The invention is useful for recombinantly producing a protein in a host-cell, by infecting or transforming a host cell capable of expressing a chelon coding sequence with a vector comprising a promoter active in the host cell operably linked to a coding region for the protein to produce a recombinant host cell and culturing the recombinant host cell under conditions, where DNA is expressed. The nucleic acid encoding the chelon protein is useful for binding divalent mercuric ions, to take up, sequester and concentrate the heavy metal ions from contaminated soil, ground water, hydroponic solutions or irrigation water of waste streams. The DNA of the invention, when immobilised onto a solid support, is useful for concentrating heavy metal ions from contaminated environment waste streams or contaminated aqueous medium including biological fluids. The nucleic acid, when recombinantly expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic), is suitable for use in the in vivo sequestration and elimination of mercuric ion from gastrointestinal tracts of animals or humans exposed to toxic metal ions such as mercury and/or cadmium. The molecules of the invention are also useful in water treatment resins. The nucleic acid of the invention is highly specific and binds divalent cation such as mercury or cadmium with high affinity. The present amino acid sequence represents one of a collection (AA097553-AA097560) of synthetic cadmium/mercury ion binding chelon proteins of the invention. This sequence is one of the heavy metal binding proteins termed chelons of the invention.

Sequence 118 AA;  
SQ

Query Match 41.0%; Score 48; DB 23; Length 118;  
Best Local Similarity 100.0%; Pred. NO. 1.2e+42;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 EHKLDVREKMDIARMETVISELVACCHARKGNVSCPSANSHPOFEK 117  
 DB 71 EHKLDVREKMDIARMETVISELVACCHARKGNVSCPSANSHPOFEK 118

## RESULT 11

ID AAY01816 standard; protein; 144 AA.

AC AAY01816;

DE 30-JUN-1999 (first entry)

DE Merr protein which has selectivity towards heavy metal ions.

KW Heavy metal ion selectivity; metal ion-specific affinity sensor;  
 KW capacitance measurement; noble metal; self-assembling monolayer;  
 KW zinc; mercury; cadmium; copper; lead; environmental sample; medicine;  
 KW food.

OS Pseudomonas aeruginosa.

PN WO9914597-A1.

PD 25-MAR-1999.

PF 15-SEP-1998; 98WO-SE01638.

PR 15-SEP-1997; 97SE-0003315.

PA (BERG/) BERGGREN C.

PA (BONT/) BONTIDEAN I.

PA (CSOE/) CSOEREGI E.

PA (JOHA/) JOHANSSON G.

PA (MATT/) MATTIASSON B.

PA (UNBI) UNIV BIRMINGHAM SCHOOL BIOLOGICAL SCI.

PA (VITO) VITO VLAAMSE INSTELLING TECHNOLOGISCH.

PI Berggren C, Bontidean I, Brown N, Corbisier P, Coerregi E;

PI Holman J, Jakeman K, Johansson G, Lloyd J, Mattiasson B;

PI Van Der Leijle D, Wilson J;

DR WPI; 1999-254424/21.

XX Capacitance sensor specific for heavy metal ions

PS Claim 8; Page 22-23; 40pp; English.

CC The present sequence represents a protein that is selective towards  
 CC heavy metal ions. It is used in the construction of the sensor of the  
 CC invention. The specification describes a metal ion-specific, affinity  
 CC sensor that measures capacitance. The sensor comprises a piece of  
 CC noble metal to which are bound groups that bind specifically to selected  
 CC heavy metal ions. These groups are bound to a self-assembling monolayer  
 CC that covers at least 90%, more preferably at least 99% of the noble  
 CC metal surface. The noble metal is a rod or piece of insulating material  
 CC (glass, quartz or silica) on which a noble metal is sputtered. The  
 CC sensor is used for qualitative or quantitative detection of selected  
 CC heavy metal ions in liquid samples, particularly of zinc, mercury,  
 CC cadmium, copper and lead in e.g. environmental samples, medicines,  
 CC foods and other products.

XX Sequence 144 AA;

Query Match 24.8%; Score 29; DB 20; Length 144;

Best Local Similarity 100.0%; Pred. No. 1.5e-22;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 GTHCEASSIAEHKLDVREKMDIARME 87  
 DB 79 GTHCEASSIAEHKLDVREKMDIARME 107

## RESULT 12

ID AAR49668 standard; Protein; 159 AA.

AC AAR49668;

DE 16-SEP-1994 (first entry)

DE Protein product of mercury resistance control gene merr(1).

KW Resistance; mercury; selectable marker; Thiobacillus ferrooxidans;  
 KW transformation; detection.

OS Thiobacillus ferrooxidans.

PN JP06000083-A.

PD 11-JAN-1994.

PF 17-JAN-1991; 91JP-0018338.

PR 17-JAN-1991; 91JP-0018338.

PA (AKIT-) AKITA KEN.

PA (DOWA) DOWA MINING CO LTD.

DR WPI; 1994-077131/10.

DR N-PSDB; AA058554.

PT Mercury resistant control gene merr and shuttle vector - for  
 PT enhanced expression of mercury resistance marker in transformed  
 PT Thiobacillus sp.

PS Disclosure; Page 2; 26pp; Japanese.

CC The mercury resistance genes can be used as selectable markers when  
 CC used to transform other bacteria.

XX Sequence 159 AA;

Query Match 20.5%; Score 24; DB 15; Length 159;

Best Local Similarity 100.0%; Pred. No. 2.9e-17;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 VLSLVACCHARKGNVSCPLIASL 54

DB 109 VLSLVACCHARKGNVSCPLIASL 132

## RESULT 13

ID AAY97662 standard; peptide; 10 AA.

AC AAY97662;

DE 08-MAY-2001 (first entry)

DE Influenza haemagglutinin peptide.

KW Bacteriophage particle; protein display;

KW Influenza haemagglutinin peptide.

OS Haemophilus influenzae.

PN WO200105950-A2.

PD 25-JAN-2001.

PF 20-JUL-2000; 2000WO-EP06968.

PR 20-JUL-1999; 99EP-0114072.

PR 18-FEB-2000; 2000EP-0103551.

XX (MORP-) MORPHOSYS AG.  
 PA  
 XX  
 PI Loehning C, Urban M, Knapplik A;  
 XX  
 DR WPI: 2001-147336/15.

XX  
 PT Displaying polypeptides on bacteriophage surface by attaching the  
 PT molecules via disulfide bonds formed between cysteine residues present  
 PT in the polypeptide and cysteine residues in the bacteriophage protein  
 PT coat -  
 XX

PS Example 2; Page 33; 80pp; English.

XX This sequence represents an influenza haemagglutinin peptide.  
 CC The invention relates to a method for displaying a polypeptide/protein on  
 CC surface of a bacteriophage particle, comprising allowing or causing  
 CC attachment of the polypeptide/protein to a unit of the protein coat of  
 CC the bacteriophage particle. The attachment is via a disulfide bond  
 CC between a cysteine residue in the polypeptide/protein and a cysteine  
 CC residue in the protein coat. Displaying polypeptide/protein on surface of  
 CC a bacteriophage particle. A collection of the bacteriophage particles  
 CC is useful for obtaining a polypeptide/protein having a desired property,  
 CC by screening the collection and/or selecting from a bacteriophage  
 CC particle to obtain a bacteriophage particle displaying a  
 CC polypeptide/protein having the desired property which is preferably  
 CC binding to a target of interest. The method further involves contacting  
 CC a collection of bacteriophage particles with the target of interest,  
 CC eluting the bacteriophage particles not binding to the target of interest  
 CC by treating the complexes of target of interest and bacteriophages under  
 CC reducing conditions. The methods easy creation and screening of large  
 CC libraries of polypeptides/proteins displayed on the surface of  
 CC bacteriophage particles.  
 XX

XX Sequence 10 AA;

XX Query Match 8.5%; Score 10; DB 22; Length 10;  
 XX Best Local Similarity 100.0%; Pred. No. 0.0012;  
 XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 SAWSHPOFEK 117  
 XX |  
 XX 1 SAWSHPOFEK 10

XX RESULT 14  
 XX AA080475  
 XX ID AA080475 standard; Peptide; 10 AA.

XX AA080475;

XX 12-MAR-2002 (first entry)

XX Peptide STREP tag.

XX Adipose protein; adp; obesity; transgenic animal; obesity;  
 XX adipositas; bulimia; wasting; cachexia; eating disorder;  
 XX body weight disorder; weight loss; cancer; infectious disease;  
 XX hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;  
 XX hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;  
 XX gastrointestinal diseases; inflammatory bowel disease;  
 XX ulcerative colitis; anorexia nervosa; glycogen storage disease;  
 XX lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;  
 XX infertility; acquired immunodeficiency syndrome; AIDS.

XX Synthetic.

XX WO200196371-A2.

XX 20-DEC-2001.

XX 13-JUN-2001; 2001WO-EP06713.

PR 16-JUN-2000; 2000US-211914P.  
 PR 23-JUN-2000; 2000EP-0113049.  
 PR 28-JUN-2000; 2000US-214518P.  
 PR 17-APR-2001; 2001EP-0109537.

XX (DEVE-) DEVELOGEN AG.

XX Breimner G, Closssek T, Dohrmann C, Haeder T, Rothe M;  
 XX WPI: 2002-106464/14.

XX Novel nucleic acid encoding adipose polypeptide which regulates, causes  
 XX or contributes to obesity, useful for treating obesity, heart disease,  
 XX hypertension, infertility, and controlling weight loss in cancer  
 XX patients -  
 XX

PS Claim 1; Page 185; 188pp; English.

XX The invention relates to a nucleic acid encoding a adipose (ADP)  
 CC polypeptide which regulates, causes or contributes to obesity in an  
 CC animal or a human. The polynucleotides, proteins, ant-adp antibodies,  
 CC modulators of adp activity, adp antisense nucleic acids, expression  
 CC vectors, adp transgenic animals are useful in the diagnosis and  
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating  
 CC disorders and/or disorders of body weight/body mass, weight loss due to  
 CC cancer or infectious diseases, genetic disorders associated with  
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,  
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,  
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative  
 CC colitis, and anorexia nervosa. They are also useful for treating  
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid  
 CC storage diseases and for treating lipomas, and/or liposarcomas. The  
 CC compositions are also useful for treating heart disease, hypertension,  
 CC and infertility and for treating conditions associated with under weight  
 CC e.g. enhancing or controlling fertility, controlling weight loss in  
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The  
 CC present sequence is a peptide tag used in a fusion protein with an  
 CC adp protein to facilitate purification  
 XX

XX Sequence 10 AA;

XX Query Match 8.5%; Score 10; DB 23; Length 10;  
 XX Best Local Similarity 100.0%; Pred. No. 0.0012;  
 XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 SAWSHPOFEK 117  
 XX |  
 XX 1 SAWSHPOFEK 10

XX RESULT 15  
 XX AAM59212  
 XX ID AAM59212 standard; peptide; 8 AA.

XX AAM59212;

XX 27-AUG-1998 (first entry)

XX Streptavidin tagged peptide ligand #2.

XX Streptavidin; ligand; binding affinity; mutant; isolation;  
 XX purification; recover; immobilise.

XX Synthetic.

XX EP835934-A2.

XX 15-APR-1998.

XX 09-OCT-1997; 97EP-0117504.

XX 10-OCT-1996; 96DE-1041876.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX Skerra A, Voss S;

XX WPI: 1998-218868/20.

DR Streptavidin mutants with higher binding affinity for peptide  
XX ligands - have mutation in amino acid region 44-53, used to isolate,  
PT purify or determine fusion proteins including these ligands

PS Claim 10; Page 11; 21pp; German.

XX AAW59211 and AAW59212 are ligands used in a method to assay binding  
CC affinity of streptavidin mutants. These mutants have a mutation within  
CC the amino acid (aa) region 44-53 of the wild-type protein show a higher  
CC binding affinity than the wild-type for peptide ligands that include the  
CC sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z  
CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin  
CC mutants can be used to isolate, purify and determine streptavidin  
CC determine/recover substances that contain streptavidin-binding groups.  
CC Such compounds may also be used to immobilise fusions on microtitre  
XX plates, microbeads or sensor chips.

SQ Sequence 8 AA;

Query Match 6.8%; Score 8; DB 19; Length 8;

Best Local Similarity 100.0%; Pred. No. 7.8e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117

DB 1 WSHPOFEK 8

Search completed: May 28, 2003, 10:39:34  
Job time : 75 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:37:17 ; Search time 44 Seconds  
(without alignments)  
255.630 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 117  
Sequence: 1 MTHCEASSLAEHLKLDVRE.....HARKGNVSCPSAMSHPOFEK 117

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	47.0	144	2	S09527 regulatory protein
2	55	47.0	144	2	S51706 regulatory protein
3	40	34.2	144	2	S51720 regulatory protein
4	29	24.8	144	2	S51755 regulatory protein
5	29	24.8	144	2	S51749 regulatory protein
6	25	21.4	151	2	S51703 regulatory protein
7	25	21.4	151	2	S51721 regulatory protein
8	25	21.4	151	2	S51705 regulatory protein
9	22	18.8	144	2	I39574 regulatory protein
10	22	18.8	144	2	S32798 merR protein - Xan
11	22	18.8	144	2	S51756 regulatory protein
12	22	18.8	144	2	S37044 regulatory protein
13	20	17.1	151	2	S51707 regulatory protein
14	15	12.8	144	2	A33858 merR protein - Bsc
15	15	12.8	144	2	A33858 merR protein - Bsc
16	15	12.8	144	2	A33858 merR protein - Bsc
17	16	12.9	129	2	T08527 trad protein - Ent
18	7	6.0	276	2	C72458 hypothetical prote
19	7	6.0	474	2	T45818 hypothetical prote
20	7	6.0	492	2	D97111 coat morphogenesis
21	7	6.0	515	2	B96825 hypothetical prote
22	7	6.0	587	2	A60367 transforming prote
23	7	6.0	1166	2	T13958 synGAP-b1 protein
24	7	6.0	1249	2	T14270 Ras-GTPase activat
25	7	6.0	1293	2	T14259 ras-GTPase-activat
26	7	6.0	4199	2	S76412 hypothetical prote
27	6	5.1	50	2	T09702 salt-inducible pro
28	6	5.1	85	2	A82035 arp synthase F0, C
29	6	5.1	98	2	T36896 hypothetical prote

30	6	5.1	107	2	B97370 hypothetical prote
31	6	5.1	107	2	AB2588 conserved hypothet
32	6	5.1	111	2	T09740 acetyl-CoA carboxy
33	6	5.1	127	2	C87640 hypothetical prote
34	6	5.1	148	2	T14784 hypothetical prote
35	6	5.1	156	2	E72518 hypothetical prote
36	6	5.1	158	2	T28955 hypothetical prote
37	6	5.1	162	2	S05712 phycoerythrin 3 alph
38	6	5.1	176	2	S10993 hypoxanthine phosph
39	6	5.1	177	2	E82306 elongation factor
40	6	5.1	185	2	D86711 probable hypoxanth
41	6	5.1	187	2	T36331 conserved hypothet
42	6	5.1	190	2	H75478 conserved hypothet
43	6	5.1	195	2	AE2145 regulatory prote
44	6	5.1	197	2	S30289 regulatory prote
45	6	5.1	198	2	G72489 hypothetical prote

## ALIGNMENTS

RESULT 1  
S09527  
regulatory protein merR - plasmid NRI  
C:Species: Plasmid NRI  
C:Date: 19-Mar-1997 #sequence-revision 29-Aug-1997 #text-change 20-Sep-1999  
C:Accession: S09527  
R:Barineau, P.; Gilbert, P.; Jackson, W.J.; Jones, C.S.; Summers, A.O.; Wisdom, S.  
J. Mol. Appl. Genet. 2, 601-619, 1984  
A:Title: The DNA sequence of the mercury resistance operon of the IncFII plasmid NRI.  
A:Reference number: S07447; MUID:85159407; PMID:6530603  
A:Accession: S09527  
A:Molecule type: DNA  
A:Residues: 1-144 <BAR>  
A:Cross-References: EMBL:R03089; NID:g150389; PIDN:AMB59072.1; PID:g455296  
C:Genetics:  
A:Genome: plasmid  
C:Superfamily: transcription repressor glr

Query Match 47.0%; Score 55; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 4.2e-50;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 THCEASSLAEHLKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOG 56  
DB 80 THCEASSLAEHLKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOG 134

RESULT 2  
S51706  
regulatory protein merR - Alcaligenes faecalis  
C:Species: Alcaligenes faecalis  
C:Date: 07-May-1995 #sequence-revision 01-Sep-1995 #text-change 20-Sep-1999  
C:Accession: S51706  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
Submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from  
A:Reference number: S51703  
A:Accession: S51706  
A:Molecule type: DNA  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <OSB>  
A:Cross-References: EMBL:Z33484; NID:g607038; PIDN:CAA83892.1; PID:g607039  
C:Superfamily: transcription repressor glr

Query Match 47.0%; Score 55; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 4.2e-50;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 THCEASSLAEHLKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOG 56  
DB 80 THCEASSLAEHLKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOG 134

RESULT 3  
S51720  
regulatory protein merr - Enterobacter aerogenes  
C:Species: Enterobacter aerogenes  
C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
C:Accession: S51720  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
A:Submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from me  
A:Reference number: S51703  
A:Accession: S51720  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <OSB>  
A:Cross-references: EMBL:Z33488; NID:9607071; PIDN:CAA83896.1; PID:9607072  
C:Superfamily: transcription repressor glnr

Query Match 34.2%; Score 40; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 2.2e-34;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 DVREKADLAMEYVLSVLCACHARKGNVSCPLIASIQ 56  
DB 95 DVREKADLAMEYVLSVLCACHARKGNVSCPLIASIQ 134

RESULT 4  
S51755  
regulatory protein merr - Pseudomonas sp.  
C:Species: Pseudomonas sp.  
C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
C:Accession: S51755  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
A:Submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from me  
A:Reference number: S51703  
A:Accession: S51755  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <OSB>  
A:Cross-references: EMBL:Z33489; NID:9607167; PIDN:CAA83897.1; PID:9607168  
C:Superfamily: transcription repressor glnr

Query Match 24.8%; Score 29; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 7.5e-23;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 GTHCEASSLAERHKLKDYREKADLARME 87  
DB 79 GTHCEASSLAERHKLKDYREKADLARME 107

RESULT 5  
S51749  
regulatory protein merr - Pseudomonas fluorescens  
C:Species: Pseudomonas fluorescens  
C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 24-May-2001  
C:Accession: S51749  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
A:Submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from me  
A:Reference number: S51703  
A:Accession: S51749  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <OSB>  
A:Cross-references: EMBL:Z33490; NID:9607153; PIDN:CAA83898.1; PID:9607154  
C:Superfamily: transcription repressor glnr

Query Match 24.8%; Score 29; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 7.5e-23;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 GTHCEASSLAERHKLKDYREKADLARME 87  
DB 79 GTHCEASSLAERHKLKDYREKADLARME 107

RESULT 6  
S51703  
regulatory protein merr - Acinetobacter calcoaceticus (isolate SE11 and SE12)  
C:Species: Acinetobacter calcoaceticus  
A:Variety: isolate SE11; isolate SE12  
C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999  
C:Accession: S51703; S51704  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
A:Submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from  
A:Reference number: S51703  
A:Accession: S51703  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <OSB>  
A:Cross-references: EMBL:Z33482; NID:9607032; PIDN:CAA83890.1; PID:9607033  
A:Experimental source: isolate SE11  
A:Accession: S51704  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <OSB>  
A:Cross-references: EMBL:Z33483; NID:9607034; PIDN:CAA83891.1; PID:9607035  
A:Experimental source: isolate SE12  
C:Genetics:  
A:Gene: merr  
C:Superfamily: transcription repressor glnr

Query Match 21.4%; Score 25; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.2e-18;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 VLSELVCACHARKGNVSCPLIASIQ 55  
DB 109 VLSELVCACHARKGNVSCPLIASIQ 133

RESULT 7  
S51721  
regulatory protein merr - Enterobacter cloacae  
C:Species: Enterobacter cloacae  
C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
C:Accession: S51721  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
A:Submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from  
A:Reference number: S51703  
A:Accession: S51721  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <OSB>  
A:Cross-references: EMBL:Z33486; NID:9607073; PIDN:CAA83894.1; PID:9607074  
C:Superfamily: transcription repressor glnr

Query Match 21.4%; Score 25; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.2e-18;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 VLSELVCACHARKGNVSCPLIASIQ 55  
DB 109 VLSELVCACHARKGNVSCPLIASIQ 133

RESULT 8  
S51705  
regulatory protein merr - Klebsiella oxytoca  
C:Species: Klebsiella oxytoca  
C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999

C:Accession: S51705  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from me  
A:Reference number: S51703  
A:Accession: S51705  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <OSB>  
A:Cross-references: EMBL:233485; NID:g607036; PIDN:CAA83893.1; PID:g607037  
C:Superfamily: transcription repressor glr

Query Match 21.4%; Score 25; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.2e-18;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 VSELYCACHARKGNVSCPLIASLQ 55  
|||||  
Db 109 VSELYCACHARKGNVSCPLIASLQ 133

RESULT 9  
139574  
mer operon regulator - Alcaligenes sp.  
C:Species: Alcaligenes sp.  
C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Sep-1999  
C:Accession: 139574; S37035  
R:Hobman, J.; Kholodil, G.; Nikiforov, V.; Ritchie, D.A.; Strike, P.; Yurleva, O.  
Gene 146; 73-78, 1994  
A:Title: The sequence of the mer operon of pMER327/419 and transposon ends of pMER327/41  
A:Reference number: 139574; MUID:94341572; PMID:8063107  
A:Accession: 139574  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-144 <RES>  
A:Cross-references: EMBL:223094; GB:L20693; NID:g388553; PIDN:AA805979.1; PID:g388554  
C:Genetics:  
A:Gene: mer  
C:Superfamily: transcription repressor glr

Query Match 18.8%; Score 22; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.6e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 LVCACHARKGNVSCPLIASLQ 56  
|||||  
Db 113 LVCACHARKGNVSCPLIASLQ 134

RESULT 10  
S32798  
mer protein - Xanthomonas sp. transposon Tn5053  
C:Species: Xanthomonas sp.  
C:Date: 08-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 20-Sep-1999  
C:Accession: S32798; S70142  
R:Kholodil, G.Y.; Yurleva, O.V.; Lomovskaya, O.L.; Gorlenko, Z.M.; Mindlin, S.Z.; Nikif  
J. Mol. Biol. 230, 1103-1107, 1993  
A:Title: Tn5053, a mercury resistance transposon with integron's ends.  
A:Reference number: S32795; MUID:93253772; PMID:8387603  
A:Accession: S32798  
A:Status: Preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-144 <KHO>  
A:Cross-references: EMBL:L03729; NID:g1019671; PIDN:AAA98396.1; PID:g154910  
A:Experimental source: plasmid RPL transposon Tn5053  
R:Kholodil, G.Y.; Mindlin, S.Z.; Bass, I.A.; Yurleva, O.V.; Minakhina, S.V.; Nikiforov,  
Mol. Microbiol. 17, 1189-1200, 1995  
A:Title: Four genes, two ends, and a res region are involved in transposition of Tn5053;  
A:Reference number: S70140; MUID:96130850; PMID:8594337  
A:Accession: S70142  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-144 <KHZ>

A:Cross-references: EMBL:L40585; NID:g710572; PIDN:AAA98322.1; PID:g710575  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995  
C:Genetics:  
A:Gene: mer  
A:Mobile element: transposon Tn5053  
C:Superfamily: transcription repressor glr

Query Match 18.8%; Score 22; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.6e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 LVCACHARKGNVSCPLIASLQ 56  
|||||  
Db 113 LVCACHARKGNVSCPLIASLQ 134

RESULT 11  
S51756  
regulatory protein merR - Comamonas testosteroni  
C:Species: Comamonas testosteroni  
C:Date: 07-May-1995 #sequence\_revision 10-Nov-1995 #text\_change 20-Sep-1999  
C:Accession: S51756  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from  
A:Reference number: S51703  
A:Accession: S51756  
A:Molecule type: DNA  
A:Residues: 1-144 <OSB>  
A:Cross-references: EMBL:233481; NID:g607169; PIDN:CAA83889.1; PID:g607170  
A:Note: the source is given as Pseudomonas testosteroni  
C:Superfamily: transcription repressor glr

Query Match 18.8%; Score 22; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.6e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 LVCACHARKGNVSCPLIASLQ 56  
|||||  
Db 113 LVCACHARKGNVSCPLIASLQ 134

RESULT 12  
S37044  
regulatory protein merR - Pseudomonas fluorescens  
N:Alternate names: mer operon regulator  
C:Species: Pseudomonas fluorescens  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Sep-1999  
C:Accession: S37044  
R:Hobman, J.; Kholodil, G.; Nikiforov, V.; Ritchie, D.A.; Strike, P.; Yurleva, O.  
submitted to the EMBL Data Library, June 1993  
A:Description: The nucleotide sequence of the mer operon of pMJ100 and transposon end  
A:Reference number: S37035  
A:Accession: S37044  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <HOB>  
A:Cross-references: EMBL:223095; NID:g397617; PIDN:CAA80641.1; PID:g397618  
C:Superfamily: transcription repressor glr

Query Match 18.8%; Score 22; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.6e-15;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 LVCACHARKGNVSCPLIASLQ 56  
|||||  
Db 113 LVCACHARKGNVSCPLIASLQ 134

RESULT 13  
S51707  
regulatory protein merR - Agrobacterium radiobacter  
C:Species: Agrobacterium radiobacter

C>Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999  
C/Accession: S51707  
R/Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
Submitted to the EMBL Data Library, May 1994  
A/Description: Sequence conservation between regulatory mercury resistance genes from me  
A/Reference number: S51703  
A/Accession: S51707  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-151 <OSB>  
A/Cross-references: EMBL:Z33487; NID:G607040; PIDN:CA83895.1; PID:G607041  
C/Superfamily: transcription repressor glrR

Query Match 17.1%; Score 20; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 2.1e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LAEHKLDVREKMDLAME 29  
|||||  
DB 88 LAEHKLDVREKMDLAME 107

## RESULT 14

A33858  
merR protein - Escherichia coli plasmid pDU1358  
C/Species: Escherichia coli  
C/Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 20-Sep-1999  
C/Accession: A33858  
R/Nucleotide: G.; Chu, L.; Silver, S.; Mista, T.K.  
J. Bacteriol. 171, 4241-4247, 1989  
A/Title: Mercury operon regulation by the merR gene of the organomercurial resistance sy  
A/Reference number: A33858; MUID:89327136; PMID:266393  
A/Accession: A33858  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-144 <NUC>  
A/Cross-references: GB:M24940; NID:G150631; PIDN:AAA98221.1; PID:G455313  
C/Genetics:  
A/Genome: Plasmid  
C/Superfamily: transcription repressor glrR  
C/Keywords: DNA binding; transcription regulation

Query Match 12.8%; Score 15; DB 2; Length 144;  
Best Local Similarity 100.0%; Pred. No. 3.5e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GTHCEASSLAEHKL 73  
|||||  
DB 79 GTHCEASSLAEHKL 93

## RESULT 15

AE0971  
50S ribosomal chain protein L28 [imported] - Salmonella enterica subsp. enterica serovar  
C/Species: Salmonella enterica subsp. enterica serovar typhi  
A/Note: this species has also been called Salmonella typhi  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 17-May-2002  
C/Accession: AE0971  
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A/Reference number: AB0502; PMID:11677608  
A/Accession: AE0971  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-78 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD03265.1; PID:G1650486; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY4066  
C/Superfamily: Escherichia coli ribosomal protein L28

Query Match 6.0%; Score 7; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 ETVLSEL 93  
|||||  
DB 65 ETVLSEL 71

Search completed: May 28, 2003, 10:42:29  
Job time : 45 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 28, 2003, 10:29:42 ; Search time 24 Seconds  
(without alignments)  
202.197 Million cell updates/sec

Title: US-09-977-137A-4

Sequence: 1 MTHCEEASSLAEHLKDYRE.....HARKGVSCPSAMSHPOFK 117

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwisProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	47.0	144	1	MERR_SALTI
2	29	24.8	144	1	MERR_PSEAE
3	15	12.8	144	1	MERR_SERMA
4	7	6.0	77	1	RL28_SALTY
5	7	6.0	129	1	TRD5_ECOLI
6	7	6.0	371	1	FIG1_RHIME
7	7	6.0	465	1	DAID_RALSO
8	7	6.0	587	1	REL_MOUSE
9	6	5.1	162	1	PHA2_PSEAE
10	6	5.1	162	1	PHA3_FREDI
11	6	5.1	176	1	HPRT_VIBNA
12	6	5.1	181	1	IF3C_GALSU
13	6	5.1	185	1	EPF_LACLA
14	6	5.1	197	1	DCRR_RHOCA
15	6	5.1	201	1	RNH2_MAGSA
16	6	5.1	203	1	FLAI_ARCFU
17	6	5.1	210	1	PRL_CARAU
18	6	5.1	210	1	PRL_CYPCA
19	6	5.1	210	1	PRL_HYPNO
20	6	5.1	211	1	LIPB_BUCAI
21	6	5.1	214	1	SH3B_MOUSE
22	6	5.1	220	1	SGBH_ECOLI
23	6	5.1	223	1	FKB3_RABIT
24	6	5.1	224	1	FKB3_BOVIN
25	6	5.1	224	1	FKB3_HUMAN
26	6	5.1	224	1	FKB3_MOUSE
27	6	5.1	241	1	AGL8_SINAL
28	6	5.1	244	1	TRUA_BACSP
29	6	5.1	253	1	VP24_MABVM
30	6	5.1	253	1	VP24_MABVP
31	6	5.1	261	1	GSHI_MOUSE
32	6	5.1	265	1	RU2A_DROME
33	6	5.1	270	1	YXEH_BACSU

34	6	5.1	284	1	SUHA_HUMAN	006520 homo sapien
35	6	5.1	291	1	YUXN_BACSU	P40950 bacillus su
36	6	5.1	293	1	GLPO_BACSU	P37965 bacillus su
37	6	5.1	304	1	MDH_HALNT	Q9HNV8 halobacteri
38	6	5.1	333	1	VACE_BACSU	P37567 bacillus su
39	6	5.1	340	1	TF2D_CAEL	P32085 caenorhabd
40	6	5.1	345	1	LEU3_LACIA	002143 lactococcus
41	6	5.1	357	1	DCUP_MYCLE	P46809 mycobacteri
42	6	5.1	357	1	REF_MYCTU	Q10605 mycobacteri
43	6	5.1	359	1	Y199_MYCTU	O07733 mycobacteri
44	6	5.1	361	1	REF_MYCLE	P45833 mycobacteri
45	6	5.1	370	1	SERC_METBA	P52878 methanosarc

## ALIGNMENTS

RESULT 1  
MERR\_SALTI  
ID MERR\_SALTI STANDARD: PRT: 144 AA.  
AC P07044;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Mercutic resistance operon regulatory protein.  
GN MERR OR HCM1.235.  
OS Salmoneilla typhi, and  
OS Shigella flexneri.  
OG Plasmid pHCM1, and Plasmid IncFII NR1.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmoneilla.  
CX NCBI\_TaxID=601, 623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhi; STRAIN=CT18; PLASMID=pHCM1;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
RA Kirogh A., Larsen T.S., Leacher S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmoneilla  
RT enterica serovar typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; PLASMID=IncFII NR1; TRANSDOSON=Tn21;  
RX MEDLINE=85159407; PubMed=6530603;  
RA Barrineau P., Gilbert P., Jackson W.J., Jones C.S., Summers A.O.,  
RA Wisdom S.;  
RT "The DNA sequence of the mercury resistance operon of the IncFII  
RT plasmid NR1.";  
RL J. Mol. Appl. Genet. 2:601-619(1984).  
CC -I- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
CC RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPERSES  
CC TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
CC WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
CC AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
CC BOUND TO THE MER SITE.  
CC -I- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; AL513383; CAD09817.1; -  
DR EMBL; K03089; AAB59072.1; -  
DR InterPro; IPR000551; HTH\_Merr.  
DR Pfam; PF00376; merr; 1.  
DR PRINTS; PR00040; HTHMERR.  
DR SMART; SM00422; HTH\_MERR; 1.  
DR PROSITE; PS00552; HTH\_MERR\_FAMILY; 1.  
DR Transcription regulation; Activator; Repressor; Mercuric resistance;  
KW Mercury; DNA-binding; Plasmid; Transposable element;  
KM Complete proteome.  
FT DNA\_BIND 10 29 H-T-H MOTIF (POTENTIAL).  
FT METAL 82 82 HG(2+).  
FT METAL 117 117 HG(2+).  
FT METAL 126 126 HG(2+).  
SQ SEQUENCE 144 AA; 15905 MW; 8BECC28A7B83EE9 CRC64;  
Query Match 47.0%; Score 55; DB 1; Length 144;  
Best Local Similarity 100.0%; Pred. No. 3.9e-51;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 THCEASSLAHKLVREKMDLARMETVLSVCACHARKGNVSCPLIASIQG 56  
DB 80 THCEASSLAHKLVREKMDLARMETVLSVCACHARKGNVSCPLIASIQG 134  
RESULT 2  
MERR\_PSEAE STANDARD; PRT; 144 AA.  
ID MERR\_PSEAE  
AC P06688;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Mercuric resistance operon regulatory protein.  
GN MERR.  
OS Pseudomonas aeruginosa, and  
OS Pseudomonas fluorescens.  
OC Plasmid pVSI.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287, 294;  
OX [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON-Tn501;  
RC MEDLINE=95014891; PubMed=6091128;  
RA Misra T.K., Brown N.L., Fritzlenger D.C., Pridmore R.D.,  
RA Barnes W.M., Haberstich L., Silver S.;  
RT "Mercuric ion-resistance operons of plasmid R100 and transposon  
Tn501: the beginning of the operon including the regulatory region  
and the first two structural genes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 81:5975-5979(1984).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=SB4;  
RC Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBD databases.  
RN [3]  
RP MOTAGENESIS.  
RP MEDLINE=90001158; PubMed=2551364;  
RA Shewchuk L.M., Verdine G.L., Nash H., Walsh C.T.;  
RT "Mutagenesis of the cysteines in the metalloregulatory protein Merr  
indicates that a metal-bridged dimer activates transcription.";  
RL Biochemistry 28:6140-6145(1989).  
CC -1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPRESSES  
TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
CC WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
CC BOUND TO THE MER SITE.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; Z00027; CAA77320.1; -  
DR EMBL; Z33489; CAA83897.1; -  
DR EMBL; Z33490; CAA83898.1; -  
DR InterPro; IPR000551; HTH\_Merr.  
DR Pfam; PF00376; merr; 1.  
DR PRINTS; PR00040; HTHMERR.  
DR SMART; SM00422; HTH\_MERR; 1.  
DR PROSITE; PS00552; HTH\_MERR\_FAMILY; 1.  
DR Transcription regulation; Activator; Repressor; Mercuric resistance;  
KW Mercury; DNA-binding; Plasmid; Transposable element;  
KM Complete proteome.  
FT DNA\_BIND 10 29 H-T-H MOTIF (POTENTIAL).  
FT METAL 82 82 HG(2+).  
FT METAL 117 117 HG(2+).  
FT METAL 126 126 HG(2+).  
FT MUTAGEN 82 82  
FT MUTAGEN 115 115  
FT MUTAGEN 117 117  
FT MUTAGEN 126 126  
FT MUTAGEN 126 126  
SQ SEQUENCE 144 AA; 15763 MW; C573298AFAF0846EF CRC64;  
Query Match 24.8%; Score 29; DB 1; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.4e-23;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 59 GTHCEASSLAHKLVREKMDLARMETVLSVCACHARKGNVSCPLIASIQG 87  
DB 79 GTHCEASSLAHKLVREKMDLARMETVLSVCACHARKGNVSCPLIASIQG 107  
RESULT 3  
MERR\_SERMA STANDARD; PRT; 144 AA.  
ID MERR\_SERMA  
AC P13111;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE Mercuric resistance operon regulatory protein.  
GN MERR.  
OS Serratia marcescens.  
OC Plasmid pD01358.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Serratia.  
OX NCBI\_TaxID=615;  
OX [1]  
RP SEQUENCE FROM N.A.  
RP MEDLINE=89327136; PubMed=2666393;  
RA Nucifora G., Chu L., Silver S., Misra T.K.;  
RT "Mercury operon regulation by the merr gene of the organomercurial  
resistance system of plasmid pD01358.";  
RT J. Bacteriol. 171:4241-4247(1989).  
CC -1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPRESSES  
TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
CC WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
CC BOUND TO THE MER SITE.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M24940: AAA98321.1; -  
 DR PIR: A33858: A33858.  
 DR InterPro: IPR000551: HTM\_Merr.  
 DR Pfam: PF00376: merr. 1.  
 DR PRINTS: PR00040: HTMERR.  
 DR SMART: SM00422: HTM\_MERR. 1.  
 DR PROSITE: PS00552: HTM\_MERR\_FAMILY. 1.  
 KW Transcription regulation; Activator; Repressor; Mercuric resistance;  
 KM Mercury; DNA-binding; Plasmid.  
 FT DNA\_BIND 10 29 H-T-H MOTIF (POTENTIAL).  
 FT METAL 82 82 HG(2+).  
 FT METAL 117 117 HG(2+).  
 FT METAL 126 126 HG(2+).  
 SO SEQUENCE 144 AA; 16033 MW; 05FBE5224B89C052 CRC64;

Query Match 12.8%; Score 15; DB 1; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 1e-08; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 GTHCEASSLAHKL 73  
 DB 79 GTHCEASSLAHKL 93

RESULT 4  
 ID RL28\_SALTY STANDARD; PRT; 77 AA.  
 AC 054325: |||||  
 DT 15-DEC-1998 (rel. 37, Created)  
 DT 15-DEC-1998 (rel. 37, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE 50S ribosomal protein L28.  
 GN RPB OR STM3728 OR STY4066.  
 OS Salmonella typhimurium, and  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602, 601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2;  
 RX MEDLINE=98055334; PubMed=9393616;  
 RA Suzuki M., Matsui K., Yamada M., Kasai H., Sofuni T., Nohmi T.;  
 RT "Construction of mutants of Salmonella typhimurium deficient in 8-  
 RT hydroxyguanine DNA glycosylase and their sensitivities to oxidative  
 RT mutagens and nitro compounds.";  
 RT Mutat. Res. 393:233-246(1997).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP SPECIES=S.typhimurium; STRAIN=LT2 / SGC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan K.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,

RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 CC -1- SIMILARITY: BELONGS TO THE L28P FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U23405: AAC01771.1; -  
 DR EMBL: AE008773: AAL22587.1; -  
 DR EMBL: AL627280: CAD03265.1; -  
 DR STYGene: SG10696: rpbm.  
 DR InterPro: IPR001383: Ribosomal\_L28.  
 DR Pfam: PF00830: Ribosomal\_L28.  
 DR TIGRFAMs: TIGR00009: L28; 1.  
 KW Ribosomal protein; Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT SEQUENCE 77 AA; 8919 MW; 58CFE6811405645 CRC64;

Query Match 6.0%; Score 7; DB 1; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 ETVLSEL 93  
 DB 64 ETVLSEL 70

RESULT 5  
 ID TRD5\_ECOLI STANDARD; PRT; 129 AA.  
 AC P27192: |||||  
 DT 01-AUG-1992 (rel. 23, Created)  
 DT 01-AUG-1992 (rel. 23, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE Trd protein.  
 GN trd.  
 OS Escherichia coli.  
 OG Plasmid IncP-beta R751.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-4.  
 RC STRAIN=HB101;  
 RX MEDLINE=92297959; PubMed=1818755;  
 RA Mele L., Strack B., Kruf V., Ianka E.;  
 RT "Gene organization and nucleotide sequence of the primase region of  
 RT IncP plasmids RP4 and R751.";  
 RL DNA Seq. 2:145-162(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Thomas C.M.;  
 RT Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO PLASMID INCP-ALPHA RP4 TRAD.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X59794: CAA2459.1; -  
 DR EMBL: U67194: AAC64471.1; -

DR PIR; S37670; S37670.  
 KW Plasmid.  
 SQ SEQUENCE 129 AA; 13600 MW; A2F7296E4463BF88 CRC64;

Query Match 6.0%; Score 7; DB 1; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 2.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 VREKAD 24  
 |||||  
 Db 32 VREKAD 38

RESULT 6  
 FLGI\_RHIME STANDARD; PRT; 371 AA.  
 ID AC 052948;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Flagellar P-ring protein precursor (Basal body P-ring protein).  
 GN FLGI OR R00664 OR SMC03032.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R011/001;  
 RA Plutzer J., Schmitt R.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RT [2]  
 RT SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Leclaire V., Masuy D.,  
 RA Pohl T., Portetelle D., Puenler A., Purnelle B., Rameberger U.,  
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
 RT Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 CC -1- FUNCTION: ASSEMBLES AROUND THE ROD TO FORM THE L-RING AND PROBABLY  
 CC PROTECTS THE MOTOR/BASAL BODY FROM SHEARING FORCES DURING  
 CC ROTATION.  
 CC -1- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE  
 CC FLAGELLAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)  
 CC MOUNTED ON A CENTRAL ROD (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE FLGI FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L49337; AAB81415.1; -  
 DR EMBL; AL591784; CAC45236.1; -  
 DR InterPro: IPR001782; Flag\_FlgI.  
 DR Pfam: PF02119; Flgi, 1.  
 DR PRINTS: PR01010; FLAGRINGFLGI.  
 KW Flagella; Periplasmic; Signal; Complete proteome.  
 FT SIGNAL 1 25  
 FT CHAIN 26 371  
 FT FLAGELLAR P-RING PROTEIN.  
 FT CONFLICT 7 7  
 FT CONFLICT 15 15  
 FT CONFLICT 323 323  
 FT SEQUENCE 371 AA; 38357 MW; 06BB4E95EE029100 CRC64;

Query Match 6.0%; Score 7; DB 1; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 KMADLAR 27  
 |||||  
 Db 237 KMADLAR 243

RESULT 7  
 DALD\_RALSO STANDARD; PRT; 465 AA.  
 ID AC P58708;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE D-arabinitol 4-dehydrogenase (EC 1.1.1.11).  
 GN DALD OR RSC2129 OR RS01475.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brotter P., Camus J.C., Catolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502(2002).  
 CC -1- CATABOLIC ACTIVITY: D-arabinitol + NAD(+) -> D-xylose + NADH.  
 CC -1- PATHWAY: D-arabinitol catabolism; first step.  
 CC -1- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL646068; CAD15836.1; -  
 DR InterPro: IPR000669; Mannitol\_ch.  
 DR Pfam: PF01232; Mannitol\_ch, 1.  
 DR PRINTS: PR00084; MTLDDHGNASE.  
 DR PROSITE: PS00974; MANNITOL\_DHGEMASE; FALSE\_NEG.  
 KW Oxidoreductase; NAD; Complete proteome.  
 SQ SEQUENCE 465 AA; 50630 MW; 47095425D83642 CRC64;

Query Match 6.0%; Score 7; DB 1; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 8.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 NVSCPSA 109  
 |||||  
 Db 203 NVSCPSA 209

RESULT 8  
 REL\_MOUSE STANDARD; PRT; 587 AA.  
 ID AC P15307;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE C-Rel proto-oncogene protein (C-Rel protein).  
 GN REL.  
 OS Mus musculus (Mouse).



```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=Spleen;
RX MEDLINE=89239482; PubMed=2654811;
RA Grumont R.J.; Gerondakis S.;
RT "Structure of a mammalian c-rel protein deduced from the nucleotide
RL sequence of murine cDNA clones.";
RN Oncogene Res. 4:1-8(1989).
RN REVISIONS
RX MEDLINE=90370373; PubMed=2204017;
RA Grumont R.J.; Gerondakis S.;
RT "The murine c-rel proto-oncogene encodes two mRNAs the expression of
RL which is modulated by lymphoid stimuli.";
CC Oncogene Res. 5:245-254(1990).
CC -1- FUNCTION: PROTO-ONCOGENE THAT MAY PLAY A ROLE IN DIFFERENTIATION
CC AND LYMPHOPROTESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X15842; CAA33843.1; ALT_SEQ.
DR EMBL; X60371; CAA42817.1; -.
DR PIR; A45505; A45505.
DR PIR; A60367; A60367.
DR HSSP; P25799; 1BFT.
DR TRANSFAC; T00169; -.
DR MGD; MGI:97897; Rel.
DR InterPro; IPR002909; IPT_TIG.
DR InterPro; IPR000451; NF_Rel_dor_fam.
DR Pfam; PF00554; RHD; 1.
DR Pfam; PF01833; TIG; 1.
DR PRINTS; PR00057; NFKB1NSCPCT.
DR SMART; SM00429; IPT; 1.
DR PROSITE; PS01204; REL_1; 1.
DR PROSITE; PS0254; REL_2; 1.
DR Kew Proto-oncogene; Phosphorylation; Nuclear protein.
FT DOMAIN 8 297
FT DOMAIN 291 296
FT MOD_RES 267 267 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 587 AA; 64960 MW; 98FC237B6D140416 CRC64;

Query Match 6.0%; Score 7; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Dubbs J.M.; Bryant D.A.;
RL Submitted (AUG-1992) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
CC FROM THE PHYCOBILIPROTEIN COMPLEX.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M99427; AAA26044.1; -.
DR HSSP; P07122; ICPC.
DR InterPro; IPR001659; Phycobilisome.
DR Pfam; PF00502; Phycobilisome; 1.
DR ProDom; PD000340; Phycobilisome; 1.
DR Phycobilisome; Electron transport; Photosynthesis; Bile pigment;
KW Multigene family.
FT BINDING 84 84 PHYCOCYANOBILIN CHROMOPHORE.
SQ SEQUENCE 162 AA; 17301 MW; 41E1840106793D6 CRC64;

Query Match 5.1%; Score 6; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 24 DLARMET 30
DB 440 DLARMET 446

RESULT 9
PHAS2_PSEA9 STANDARD; PRT; 162 AA.
AC 052452;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE C-phycocyanin-2 alpha chain.
GN CPC2A
OS Pseudanabaena sp. (strain PCC 7409).
OC Bacteria; Cyanobacteria; Oscillatoriales; Pseudanabaena.
OX NCBI_TaxID=29415;

```

```

OY 7 ASSLA 12
DB 48 ASSLA 53

RESULT 10
PHAS3_FREDI STANDARD; PRT; 162 AA.
AC P14876;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE C-phycocyanin-3 alpha chain.
GN CPC3
OS Fremyella diplosiphon (Calothrix PCC 7601).
OC Bacteria; Cyanobacteria; Nostocales; Rivulariaceae; Fremyella.
OX NCBI_TaxID=1197;
RN [1]
RP SEQUENCE FROM N.A.
RA Mazel D.; Hounard J.; Tandeau de Marsac N.;
RT "A multigene family in Calothrix sp. PCC 7601 encodes phycocyanin,
RL the major component of the cyanobacterial light harvesting antenna.";
RN Mol. Gen. Genet. 211:296-304(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89384903; PubMed=2506452;
RA Mazel D.; Marliere P.;
RT "Adaptive eradication of methionine and cysteine from cyanobacterial
RL light-harvesting proteins.";
RN Nature 341:245-248(1989).
CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
CC FROM THE PHYCOBILIPROTEIN COMPLEX.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- INDUCTION: PHYCOCYANIN-3 IS EXPRESSED IN RED LIGHT UNDER
CC CONDITIONS OF SULFUR DEPRIVATION.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR EMBL: X06083; CAA29460.1; -

DR PIR: S05712; S05712.

DR HSSP: P07122; ICPC.

DR InterPro: IPR001659; Phycobilisome.

DR Pfam: PF00502; Phycobilisome; 1.

DR ProDom: PD000340; Phycobilisome; 1.

DR Phycobilisome; Electron transport; Photosynthesis; Bile pigment;

DR Multigene family.

FT BINDING 84 PHYCOCYANOBILIN CHROMOPHORE.

SO SEQUENCE 162 AA; 17392 MW; 0679FC970554EDD6 CRC64;

QY Query Match 5.1%; Score 6; DB 1; Length 162;

DB Best Local Similarity 100.0%; Pred. No. 39;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ASLAE 12

DB 48 ASLAE 53

RESULT 11

HPRT\_VIBHA STANDARD; PRT; 176 AA.

AC P18134;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypoxanthine phosphoribosyltransferase (EC 2.4.2.8) (HPRT).

GN HPT.

OS Vibrio harveyi.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=669;

RN [1]

RE SEQUENCE FROM N.A.

RA MEDLINE:90356428; PubMed-2388850;

RT Showalter R.E., Silverman M.R.;

RT "Nucleotide sequence of a gene, hpt, for hypoxanthine

phosphoribosyltransferase from *Vibrio harveyi*."

RL Nucleic Acids Res. 18:4621-4621(1990).

CC -1- FUNCTION: THIS ENZYME ACTS EXCLUSIVELY HYPOXANTHINE; IT DOES NOT

ACT ON GUANINE.

CC -1- CATALYTIC ACTIVITY: IMP + diphosphate = hypoxanthine + 5-phospho-

alpha-D-ribose 1-diphosphate.

CC -1- PATHWAY: Purine salvage.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE

PHOSPHORIBOSYLTRANSFERASE FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: X53382; CAA37462.1; -

DR PIR: S10993; S10993.

DR HSSP: Q26997; IQK3.

DR InterPro: IPR000836; PRTtransferase.

DR InterPro: IPR002375; Pr/PY-TP-transf.

DR Pfam: PF00156; Pribosyltran; 1.

DR TIGRFAms: TIGR01203; HGPRTase; 1.

DR PROSITE: PS00103; PUR-PYR-TRANSFER; 1.

KW Transferase; Glycosyltransferase; Purine salvage; Magnesium.

FT METAL 98 98 MAGNESIUM (BY SIMILARITY).

FT METAL 99 99 MAGNESIUM (BY SIMILARITY).

SO SEQUENCE 176 AA; 19963 MW; 2027D73A9CAANCE CRC64;

Query Match 5.1%; Score 6; DB 1; Length 176;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 MADLAR 27

DB 48 MADLAR 53

RESULT 12

IF3C\_GALSU STANDARD; PRT; 181 AA.

AC Q9MS97;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Translation Initiation Factor IF-3, chloroplast.

GN INFC OR IF3.

OS Gallieria sulphuraria.

OC Chloroplast.

OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;

OC Gallieria.

OX NCBI\_TaxID=130081;

RN [1]

RE SEQUENCE FROM N.A.

RA STRAIN-UTEX 2393;

RT Whitney S.M., Andrews J.;

RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE

ECUILLIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN

FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S

SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.

CC -1- SUBUNIT: MONOMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Chloroplast.

CC -1- SIMILARITY: BELONGS TO THE IF-3 FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: AF233069; AAF81685.1; -

DR HSSP: P03000; ITRF.

DR InterPro: IPR001288; IF3.

DR Pfam: PF00707; IF3; 1.

DR ProDom: PD002880; IF3; 1.

DR TIGRFAms: TIGR00168; INFC; 1.

DR PROSITE: PS00938; IF3; 1.

KW Initiation factor; Protein biosynthesis; Chloroplast.

SO SEQUENCE 181 AA; 21393 MW; 572014236DE7ABF8 CRC64;

Query Match 5.1%; Score 6; DB 1; Length 181;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 SSLAEH 71

DB 156 SSLAEH 161

RESULT 13

EFP\_LACLA STANDARD; PRT; 185 AA.

AC Q9CHN6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Elongation factor P (EF-P).

GN EFP OR L10692.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.

```

OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Mincer P., Manger S., Sorokin A.;
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RL lactis ssp. lactis IL1403."
CC -1- FUNCTION: INVOLVED IN PEPTIDE BOND SYNTHESIS. STIMULATES EFFICIENT
CC TRANSLATION AND PEPTIDE BOND SYNTHESIS ON NATIVE OR RECONSTITUTED
CC 70S RIBOSOMES IN VITRO. PROBABLY FUNCTIONS INDIRECTLY BY ALTERING
CC THE AFFINITY OF THE RIBOSOME FOR AMINOACYL-TRNA, THUS INCREASING
CC THEIR REACTIVITY AS ACCEPTORS FOR PEPTIDYL TRANSFERASE (BY
CC SIMILARITY).
CC -1- PATHWAY: Protein biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ELONGATION FACTOR P FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006302; AAK04790.1; -
DR InterPro; IPR001059; EF-P.
DR Pfam; PF01132; EFP; 1.
DR TIGRPFAMS; TIGR00038; efp; 1.
DR PROSITE; PS01275; EFP; 1.
DR Protein biosynthesis; Elongation factor; Complete proteome.
SQ SEQUENCE 185 AA: 20653 MW: B97694B241DCFD52 CRC64;

Query Match 5.1%; Score 6; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 KLDYR 19
Db 40 KLDYR 45

RESULT 14
DCTR_RHOCA STANDARD; PRT; 197 AA.
ID DCTR_RHOCA
AC P37740;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE C4-dicarboxylate transport transcriptional regulatory protein dctr.
GN DCTR.
OS Rhodospirillum rubrum (Rhodospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum;
OC Rhodospirillum;
OC NCBI_TaxID=1061;
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93204897; PubMed=8455557;
RA Hamblin M.J., Shaw J.G., Kelly D.J.;
RT "Sequence analysis and interposon mutagenesis of a sensor-kinase
RT (Dcts) and response-regulator (Dctr) controlling synthesis of the
RT high-affinity C4-dicarboxylate transport system in Rhodospirillum
RL capsulatus."
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR
CC INVOLVED IN THE TRANSPORT OF C4-DICARBOXYLATES. DCTR FUNCTIONS AS
CC A TRANSCRIPTIONAL REPRESSOR OF GENES FOR C4-DICARBOXYLATE
CC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

```

```

CC -1- PTM: PHOSPHORYLATED BY DCTS.
CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64733; CAA46000.1; -
DR PIR; S30289; S30289.
DR HSSP; P10958; IDBW.
DR InterPro; IPR000792; HTH_LuxR.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00072; response_reg; 1.
DR Pfam; PF00196; GcE; 1.
DR PRINTS; PR00038; HTH_LuxR.
DR PRODOM; PD000039; Response_reg; 1.
DR PRODOM; PD000307; HTH_LuxR; 1.
DR SMART; SM00421; HTH_LuxR; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS00622; HTH_LuxR_FAMILY; 1.
DR PROSITE; PS0110; RESPONSE_REGULATORY; 1.
KW Sensory transduction; Phosphorylation; Transcription regulation;
KW Activator; DNA-binding.
FT DOMAIN 1 120 RESPONSE REGULATORY.
FT MOD_RES 53 53 PHOSPHORYLATION (BY SIMILARITY).
FT DOMAIN 128 143 INTER-DOMAIN LINKER (POTENTIAL).
FT DNA_BIND 160 179 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 197 AA: 21270 MW: C5F72003FCACAD4 CRC64;

Query Match 5.1%; Score 6; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 ADLARM 86
Db 191 ADLARM 196

RESULT 15
RNH2_MAGSA STANDARD; PRT; 201 AA.
ID RNH2_MAGSA
AC Q50412;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease HII (Ec 3.1.26.4) (RNase HII).
GN RNH.
OS Magnetospirillum rubrum (Magnetospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum;
OC Magnetospirillum;
OC NCBI_TaxID=84159;
OX NCBI_TaxID=84159;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-AMB-1;
RA Matsunaga T.;
RT Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THIS ENZYME IS AN ENDONUCLEASE THAT DEGRADES THE RNA OF
CC RNA-DNA HYBRIDS SPECIFICALLY (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphonooxyster.
CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE RNASE HII FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; D32253; BAA06983.1; -.  
DR HSSP; Q57599; IEKE.  
DR InterPro; IPR001352; RNase\_HII/HIII.  
DR Pfam; PF01351; RNase\_HII; 1.  
KW Hydrolase; Nuclease; Endonuclease; Manganese.  
FT ACT\_SITE 18 18 BY SIMILARITY.  
FT ACT\_SITE 113 113 BY SIMILARITY.  
FT ACT\_SITE 131 131 BY SIMILARITY.  
SQ SEQUENCE 201 AA; 21597 MW; 0D01B23CF85DD2A2 CRC64;

Query Match 5.1%; Score 6; DB 1; Length 201;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 22 MADLAR 27  
|||||  
Db 152 MADLAR 157

Search completed: May 28, 2003, 10:40:07  
Job time : 26 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:35:28 ; Search time 84 Seconds  
(without alignments)

286.994 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 117  
Sequence: 1 THCEASSLAEHKLKDVRE.....HARKGVSCPSAMSHPOFEK 117

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL.21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioplasmid:\*  
17: sp\_archaeoplasmid:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	47.0	124	2	08RRK29 pseudomonas
2	55	47.0	144	2	044190 alcaligenes
3	40	34.2	144	2	046655 enterobacte
4	35	29.9	111	2	09R9X0 pseudomonas
5	29	24.8	144	2	052395 pseudomonas
6	29	24.8	172	2	09AFK4 shigella fl
7	25	21.4	151	2	057492 enterobacte
8	25	21.4	151	2	007304 pseudomonas
9	25	21.4	151	2	057106 acinetobact
10	22	18.8	110	2	09R9W9 pseudomonas
11	22	18.8	144	2	060233 alcaligenes
12	21	17.9	151	2	0934S8 thiothacilli
13	20	17.1	151	2	044191 agrobacteri
14	20	17.1	151	2	099093 acinetobact
15	20	17.1	151	16	0935L5 salmonella
16	15	12.8	144	2	007300 pseudomonas

17	15	12.8	144	2	P77071 escherichia
18	13	11.1	141	2	054316 thiothacilli
19	13	11.1	151	2	099092 acinetobact
20	7	6.0	257	17	097500 sulfolobus
21	7	6.0	276	17	09Y9H7 aeropyrum p
22	7	6.0	312	17	09HH01 pyrococcus
23	7	6.0	363	11	09D511 mus musculus
24	7	6.0	474	10	09M2K7 arabidopsis
25	7	6.0	476	10	08M4L8 arabidopsis
26	7	6.0	492	16	0971D5 clostridium
27	7	6.0	515	10	09SAK7 qesak7 arabidopsis
28	7	6.0	902	4	08TCS2 qesak7 arabidopsis
29	7	6.0	1122	5	09VCR8 qesak7 arabidopsis
30	7	6.0	1166	11	09ETR1 qesak7 arabidopsis
31	7	6.0	1171	11	09QX12 qesak7 arabidopsis
32	7	6.0	1179	4	09GE2 qesak7 arabidopsis
33	7	6.0	1249	11	09QX02 qesak7 arabidopsis
34	7	6.0	1284	11	09ESK6 qesak7 arabidopsis
35	7	6.0	1293	11	09Q0H6 qesak7 arabidopsis
36	7	6.0	2659	5	09VBU7 qesak7 arabidopsis
37	7	6.0	4199	16	P74440 synecocyst
38	7	6.0	8817	2	033840 polydium
39	6	5.1	50	10	004137 medicago sa
40	6	5.1	53	6	P82126 sus scrofa
41	6	5.1	81	15	069007 human endog
42	6	5.1	84	2	030342 vibrio chol
43	6	5.1	85	16	09KMH0 vibrio chol
44	6	5.1	86	6	019129 bos taurus
45	6	5.1	97	8	08WH74 fraseria tub

#### ALIGNMENTS

RESULT 1					
08RRK29	PRELIMINARY;	PRT;	124 AA.		
AC 08RRK29:					
DT 01-JUN-2002 (TREMBL)	21, Created				
DT 01-JUN-2002 (TREMBL)	21, Last sequence update				
DT 01-JUN-2002 (TREMBL)	21, Last annotation update				
DE Merr protein (Fragment).					
GN MERR.					
OS Pseudomonas fluorescens.					
OC Plasmid pKH22.					
CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;					
OC Pseudomonas.					
OX NCBI_TaxID=294;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TRANSPOSON-TN21 HOMOLOGUE;					
RA Kholodil G.Y., Gorlenko Z.M., Mindlin S.Z., Nikiforov V.G.;					
RT "Distribution of distinct microvariants of tm5041 in environmental bacteria."					
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.					
DR EMBL: AJ422225; CAD19592.1; -					
KW Plasmid.					
FT NON-TER					
SO SEQUENCE	124 AA;	13759 MW;	C514222BB0ED7754 CRC64;		
QY	2	THCEASSLAEHKLKDVREKMDLAEVYLSLVACARKNVSCPLASLOG	56		
DB	60	THCEASSLAEHKLKDVREKMDLAEVYLSLVACARKNVSCPLASLOG	114		
RESULT 2					
044190	PRELIMINARY;	PRT;	144 AA.		
AC 044190:					

DR 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Regulatory protein.  
 GN MERR.  
 OS Alcaligenes faecalis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 CC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 RX NCBL\_TaxID=511;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Osborn A.M., Bruce K.D., Ritchie D.A.;  
 RT "Sequence Conservation between Regulatory Mercury Resistance Genes  
 from Mercury Polluted and Pristine Environments."  
 RT Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 CC EMBL; Z33484; CAAB3892.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PR00040; HTHMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 DR DNA-binding; transcription regulation.  
 KW SEQUENCE 144 AA; 15832 MW; B71DFF3C980DC49 CRC64;  
 SQ

Query Match 47.0%; Score 55; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 2e-50;  
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THREEASLAEHKLIDYREKADLARMETVSELVACACHARKGNVSCPLIASLOG 56  
 DB 80 THREEASLAEHKLIDYREKADLARMETVSELVACACHARKGNVSCPLIASLOG 134

## RESULT 3

ID 046655 PRELIMINARY; PRT; 144 AA.  
 AC 046655;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Regulatory protein.  
 GN MERR.  
 OS Enterobacter aerogenes (aerobacter aerogenes).  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Enterobacter.  
 RX NCBL\_TaxID=548;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Osborn A.M., Bruce K.D., Ritchie D.A.;  
 RT "Sequence Conservation between Regulatory Mercury Resistance Genes  
 from Mercury Polluted and Pristine Environments."  
 RT Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 CC EMBL; Z33488; CAAB3896.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PR00040; HTHMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 DR DNA-binding; transcription regulation.  
 KW SEQUENCE 144 AA; 15652 MW; 1D6E1F50D37A1337 CRC64;  
 SQ

Query Match 34.2%; Score 40; DB 2; Length 144;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-34;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 17 DYREKADLARMETVSELVACACHARKGNVSCPLIASLOG 56  
 ||||||||||||||||||||||||||||||||||||||||

DB 95 DYREKADLARMETVSELVACACHARKGNVSCPLIASLOG 134

## RESULT 4

ID 09R9X0 PRELIMINARY; PRT; 111 AA.  
 AC 09R9X0;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE Merr (fragment).  
 GN MERR.  
 OS Pseudomonas putida.  
 OC Plasmid group 2 plasmid.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 RX NCBL\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bruce K.D., Lilley A.K., Bailey M.J.;  
 RT "mer sequences on plasmids."  
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF092069; AAD52705.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR Plasmid.  
 KW NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 111 AA; 12406 MW; 642A1FD89A898C97 CRC64;

Query Match 29.9%; Score 35; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-29;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 DYREKADLARMETVSELVACACHARKGNVSCPLI 51  
 DB 72 DYREKADLARMETVSELVACACHARKGNVSCPLI 106

## RESULT 5

ID 052395 PRELIMINARY; PRT; 144 AA.  
 AC 052395;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Putative Merr protein (Organomercurial resistance regulatory  
 protein).  
 GN MERR.  
 OS Pseudomonas putida, and  
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
 OC Plasmid pMW0, and Plasmid pPB.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 RX NCBL\_TaxID=303, 316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SPECIES-P.putida; PLASMID-PMW0;  
 RA "Created A., Lamberton L., Williams P.A., Thomas C.M.;  
 RT "Complete nucleotide sequence of Incp-9 plasmid pMW0."  
 RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC [2]  
 RP SEQUENCE FROM N.A.  
 RA SPECIES-P.putida; STRAIN-OX; PLASMID-PPB;  
 RX MEDLINE-85014891; PubMed-6091128;  
 RA Mista T.K., Brown N.L., Filtzinger D.C., Pridmore R.D., Barnes W.M.,  
 RT "Mercuric ion-resistance operons of plasmid R100 and transposon Tn501:  
 RT the beginning of the operon including the regulatory region and the  
 RT first two structural genes."  
 RN Proc. Natl. Acad. Sci. U.S.A. 81:5975-5979(1984).  
 [3]

```
RP SEQUENCE FROM N.A.
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;
RX MEDLINE=86174347; PubMed=3007931;
RA Brown N.L., Misra T.K., Winnie J.N., Schmidt A., Seiff M., Silver S.;
RT "The nucleotide sequence of the mercury resistance operons of plasmid
RT R100 and transposon Tn501: further evidence for mer genes which
RT enhance the activity of the mercury ion detoxification system.";
RL Mol. Gen. Genet. 202;143-151(1986).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;
RX MEDLINE=93253772; PubMed=8387603;
RA Khlodil G.Ya., Yurleva O.V., Lomovskaya O.L., Gorlenko Zh.M.,
RA Mindlin S.Z., Nikiforov V.G.;
RT "Tn5053, a mercury resistance transposon with integron's ends.";
RL J. Mol. Biol. 230;1103-1107(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;
RX MEDLINE=96105204; PubMed=8529897;
RA Rentero D., Gall E., Barbieri P.;
RT "Cloning and comparison of mercury- and organomercurial-resistance
RT determinants from a Pseudomonas stutzeri plasmid.";
RL Gene 166;77-82(1995).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;
RX MEDLINE=96148002; PubMed=9479042;
RA Rentero D., Mozzon E., Gall E., Barbieri P.;
RT "Two aberrant mercury resistance transposons in the Pseudomonas
RT stutzeri plasmid PPB.";
RL Gene 208;37-42(1998).
CC -I- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AJ344068; CAC66844.1; -.
DR EMBL; U90263; AAC8229.1; -.
DR InterPro; IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PRO0040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW Plasmid; DNA-binding; transcription regulation.
SQ SEQUENCE 144 AA; 15884 MW; F5760BEC8602FC7 CRC64;

Query Match 24.8%; Score 29; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 7.9e-23;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GTHCEASSLAHKLDVREKMDLARME 87
DB 79 GTHCEASSLAHKLDVREKMDLARME 107

RESULT 6
QYAFK4 PRELIMINARY; PRT; 172 AA.
AC QYAFK4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Tn501 repressor.
GN MERR.
OS Shigella flexneri.
OC Plasmid virulence plasmid pWR501.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21189246; PubMed=11292750;
RA Venkatesan M.M., Goldberg W.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blatner F.R.;
RT "Complete DNA sequence and analysis of the large virulence plasmid of
```

```
RT Shigella flexneri.";
RL Infect. Immun. 69;3271-3285(2001).
CC -I- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF348706; AAK18578.1; -.
DR InterPro; IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PRO0040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW DNA-binding; Plasmid; transcription regulation.
SQ SEQUENCE 172 AA; 18826 MW; 897D139E7BC182A9 CRC64;

Query Match 24.8%; Score 29; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 9.3e-23;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 GTHCEASSLAHKLDVREKMDLARME 87
DB 107 GTHCEASSLAHKLDVREKMDLARME 135

RESULT 7
QYAFK2 PRELIMINARY; PRT; 151 AA.
ID QYAFK2;
AC QYAFK2;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Regulatory protein.
GN MERR.
OS Enterobacter cloacae, and
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550, 571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SO1;
RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBD databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SE31;
RA Osborn A.M.;
RL Submitted (MAY-1994) to the EMBL/Genbank/DBD databases.
CC -I- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; Z33486; CAA83894.1; -.
DR EMBL; Z33485; CAA83893.1; -.
DR InterPro; IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PRO0040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 151 AA; 16559 MW; 238460FCE51754AD CRC64;

Query Match 21.4%; Score 25; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 VLSLVACAHARKGNVCSPLIASIQ 55
DB 109 VLSLVACAHARKGNVCSPLIASIQ 133

RESULT 8
QYAFK4 PRELIMINARY; PRT; 151 AA.
ID QYAFK4;
AC QYAFK4;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
```

DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE MERR protein (Mercuric resistance operon regulatory protein).  
 MERR.  
 OS Pseudomonas sp.  
 OC Bacteria: Proteobacteria.  
 OX NCBI\_TaxID=306;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TC97;  
 RX MEDLINE=97303088; PubMed=9159519;  
 RA Yurleva O., Kholodil G., Minakhin L., Gorlenko Z., Kalyaeva E.,  
 RA Mindlin S., Nikiforov V.;  
 RT "Intercontinental spread of promiscuous mercury-resistance transposons  
 in environmental bacteria.";  
 RL Mol. Microbiol. 24:321-329(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TC97;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Kholodil G.Y., Mindlin S.Z., Gorlenko Z.M., Bass I.A., Kalyaeva E.S.,  
 RA Nikiforov V.;  
 RT "Molecular genetic analysis of the Tn5041 transposition system";  
 RL Russ. J. Genet. 36:365-373(2000).  
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 DR EMBL: Y09210; CAA70409.2; -;  
 DR EMBL: Y18976; CAB81570.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PRO0040; HTMMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 KW DNA-binding; Transcription regulation.  
 SQ SEQUENCE 151 AA; 16541 MW; 17CC8F1005A33FDD CRC64;  
 QY Query Match 21.4%; Score 25; DB 2; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 109 VLSEVACACHARKGNVSCPLIASIQ 133  
 QY 31 VLSEVACACHARKGNVSCPLIASIQ 55  
 ||||||||||||||||||||  
 ID 057106 PRELIMINARY; PRT; 151 AA.  
 AC 057106; 008282; 008130; 008287; 008166; 008288; 008185;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Regulatory protein (Mercury resistance operon regulatory protein) (MRR  
 operon regulatory protein).  
 GN MERR.  
 OS Acinetobacter calcoaceticus,  
 OS Escherichia coli,  
 OS Alcaligenes sp.,  
 OS Pantoea agglomerans,  
 OS Enterobacter cloacae,  
 OS Acinetobacter sp., and  
 OS Acinetobacter sp. LS56-7.  
 OC Bacteria: Proteobacteria; gamma subdivision; Moraxellaceae;  
 OC Acinetobacter.  
 OX NCBI\_TaxID=471, 562, 512, 549, 550, 472, 107402;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SE12;  
 RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CH210; TRANSPOSON-TN5059;  
 RX MEDLINE=97303088; PubMed=9159519;  
 RA Yurleva O., Kholodil G., Minakhin L., Gorlenko Z., Kalyaeva E.,  
 RA Mindlin S., Nikiforov V.;  
 RT "Intercontinental spread of promiscuous mercury-resistance transposons  
 in environmental bacteria.";  
 RL Mol. Microbiol. 24:321-329(1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Acinetobacter sp.; STRAIN=BM3; PLASMID=PKLH207;  
 RC TRANSPOSON-TNPKLH207 TNPKLH2-LIKE ABERRANT TRANSPOSON;  
 RA Kholodil G.Y., Yurleva O.V., Mindlin S.Z., Gorlenko Z.M.,  
 RA Nikiforov V.G.;  
 RT "PKLH2-like aberrant transposons and possible mechanisms of their  
 dissemination.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Acinetobacter sp. LS56-7; PLASMID=PKLH204;  
 RC TRANSPOSON-TNPKLH204;  
 RX MEDLINE=21272500; PubMed=11376944;  
 RA Kholodil G.Y.;  
 RT "The shuffling function of resolvases.";  
 RL Gene 269:121-130(2001).  
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 DR EMBL: Y09026; CAA70240.1; -;  
 DR EMBL: Y08993; CAA70195.1; -;  
 DR EMBL: Z33483; CAA83891.1; -;  
 DR EMBL: Z33482; CAA83890.1; -;  
 DR EMBL: Y08992; CAA70185.1; -;  
 DR EMBL: Y09025; CAA70237.1; -;  
 DR EMBL: AJ245842; CAC80722.1; -;  
 DR EMBL: AJ250860; CAC38823.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PRO0040; HTMMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 KW DNA-binding; Plasmid; Transcription regulation.  
 SQ SEQUENCE 151 AA; 16529 MW; 239350FCE51754AD CRC64;  
 QY Query Match 21.4%; Score 25; DB 2; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-18;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 109 VLSEVACACHARKGNVSCPLIASIQ 133  
 QY 31 VLSEVACACHARKGNVSCPLIASIQ 55  
 ||||||||||||||||||||  
 ID 09R9W9 PRELIMINARY; PRT; 110 AA.  
 AC 09R9W9;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE Merr (Fragment).  
 GN MERR.  
 OS Pseudomonas putida.  
 OS Plasmid group 5 plasmid.  
 OC Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=303;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KT2440;  
 RA Bruce K.D., Lilley A.K., Bailey M.J.;  
 RT "mer sequences on plasmids";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF092070; AAD52706.1; -;



```

DR InterPro: IPR000551; HTH_MERR.
DR SMART: SMO0422; HTH_MERR; 1.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 110 110
SQ SEQUENCE 110 AA: 12211 MW: 2C7C09EB8ACB7BCA CRC64:
Query Match 18.8%; Score 22; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 1.7e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 LVCACHARKGNVSCPLIASLOG 56
DB 80 LVCACHARKGNVSCPLIASLOG 101
RESULT 11
060233 PRELIMINARY: PRT: 144 AA.
ID 060233
AC 060233; P75015; Q44314; Q51768; Q51808; Q52601; Q56390; Q56443;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Mercuric resistance operon regulatory protein (MERR protein).
GN MERR.
OS Alcaligenes sp..
OS Pseudomonas fluorescens,
OS Comamonas testosteroni (Pseudomonas testosteroni),
OS Xanthomonas, and
OS unidentified.
OC Pseudomonas sp. (strain ADP).
OC Plasmid pMER327, Plasmid Rpl. and Plasmid PADP-1.
OC Bacteria: Proteobacteria; beta subdivision; Alcaligenaceae;
OC Alcaligenes.
OC NCBI_TaxID=512, 294, 285, 32644, 338, 47660;
OX [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Alcaligenes sp., and P.fluorescens; TRANSPOSON=TN5053;
RX MEDLINE=9431572; Pubmed=8063107;
RA Hobman J., Kholodil G., Nikiforov V., Ritchie D.A., Strike P.,
RA Yurleva O.;
RT "The sequence of the mer operon of pMER327/419 and transposon ends of
RT pMER327/419, 330 and 05."
RL Gene 146:73-78(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.testosteroni; STRAIN=SE3;
RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=unidentified;
RX MEDLINE=96130850; Pubmed=8594337;
RA Kholodil G.Y., Mindlin S.Z., Bass I.A., Yurleva O.V., Minakhina S.V.,
RA Nikiforov V.G.;
RT "Four genes, two ends, and a res region are involved in transposition
RT of TN5053: a paradigm for a novel family of transposons carrying
RT either a mer operon or an integron."
RL Mol. Microbiol. 17:1189-1200(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=unidentified;
RA Kholodil G.Y.;
RL Russ. J. Genet. 31:1447-1451(1995).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=Xanthomonas; PLASMID=RPI; TRANSPOSON=TN5053;
RX MEDLINE=93253772; Pubmed=8387603;
RA Kholodil G.Ya., Yurleva O.V., Lomovskaya O.L., Gorlenko Zh.M.,
RA Mindlin S.Z., Nikiforov V.G.;
RT "TN5053, a mercury resistance transposon with integron's ends."
RL J. Mol. Biol. 230:1103-1107(1993).
RN [6]

```

```

RP SEQUENCE FROM N.A.
RC SPECIES=Pseudomonas sp. (strain ADP); STRAIN=ADP; PLASMID=PADP-1;
RA Martinez B.M., Tomkins J., Wackett L.P., Wing R., Sadowsky M.J.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL: X73112; CAA51538.1; -
DR EMBL: L20693; AAB05979.1; -
DR EMBL: Z23095; CAA80641.1; -
DR EMBL: L20694; AAB02644.1; -
DR EMBL: Z23094; CAA80640.1; -
DR EMBL: L40585; AAA98322.1; -
DR EMBL: Z33481; CAA83889.1; -
DR EMBL: U66917; AAK50289.1; -
DR InterPro: IPR000551; HTH_MERR.
DR Pfam: PF00376; merr; 1.
DR PRINTS: PR00040; HTHMERR.
DR SMART: SMO0422; HTH_MERR; 1.
DR PROSITE: PS00552; HTH_MERR_FAMILY; 1.
KW Activator; DNA-binding; Mercuric resistance; Mercury; Plasmid;
KW Repressor; Transcription regulation.
SQ SEQUENCE 144 AA: 16060 MW: E4B3EFEECA317F2D CRC64:
Query Match 18.8%; Score 22; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.1e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 LVCACHARKGNVSCPLIASLOG 56
DB 113 LVCACHARKGNVSCPLIASLOG 134
RESULT 12
093458 PRELIMINARY: PRT: 151 AA.
ID 093458
AC 093458;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Mer operon regulatory protein.
GN MERR.
OS Thiobacillus ferrooxidans.
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.
OC NCBI_TaxID=920;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G66; TRANSPOSON=TN5037;
RA Kalyeava E.S., Kholodil G.Y., Bass I.A., Gorlenko A.M., Yurleva O.V.,
RA Nikiforov V.G.;
RT "TN5037, a Tn21-like mercury resistance transposon from Thiobacillus
RT ferrooxidans."
RL Russ. J. Genet. 37:972-975(2001).
CC -1 SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL: AJ251743; CAC69248.1; -
DR InterPro: IPR000551; HTH_MERR.
DR Pfam: PF00376; merr; 1.
DR PROSITE: PS00552; HTH_MERR_FAMILY; UNKNOWN.1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 151 AA: 16555 MW: 26DDF7A510B0829B CRC64:
Query Match 17.9%; Score 21; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 LVCACHARKGNVSCPLIASLOG 55
DB 113 LVCACHARKGNVSCPLIASLOG 133
RESULT 13
044191 PRELIMINARY: PRT: 151 AA.
ID 044191

```

AC Q44191;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Regulatory protein.  
GN MERR.  
OS Agrobacterium tumefaciens.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_Taxid=358;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T217;  
RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
RT "Sequence Conservation between Regulatory Mercury Resistance Genes  
RT from Mercury Polluted and Pristine Environments.";  
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL: Z33487; CAA83895.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr.1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR.1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
KW DNA-binding; transcription regulation.  
SQ SEQUENCE 151 AA; 16591 MW; 548460FCE50240FC CRC64;  
Query Match 17.1%; Score 20; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 2, 9e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 LAEHLKDVREKMDLAME 29  
DB 88 LAEHLKDVREKMDLAME 107  
RESULT 14  
ID Q99093 PRELIMINARY; PRT; 151 AA.  
AC Q99093;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Mer operon regulatory protein.  
GN MERR.  
OS Acinetobacter calcoaceticus,  
OS Acinetobacter lwoffi, and  
OS Acinetobacter sp.  
OC Plasmid pKLIH2, Plasmid pKLIH3, Plasmid pKLIH102, Plasmid pKLIH104, and  
OC Plasmid pKLIH205.  
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
OC Acinetobacter.  
OX NCBI\_Taxid=471, 28090, 472;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.calcoaceticus; PLASMID=pKLIH2;  
RA Lomovskaya O.L., Nikiforov V.G.;  
RA Genetika 24:1064-1071(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.calcoaceticus; PLASMID=pKLIH2;  
RX MEDLINE-94134837; PubMed-8302940;  
RA Kholodil G.Y., Lomovskaya O.L., Gorlenko Z.M., Mindlin S.Z.,  
RA Yurleva O.V., Nikiforov V.G.;  
RT "Molecular characterization of an aberrant mercury resistance  
RT transposable element from an environmental Acinetobacter strain.";  
RL Plasmid 30:303-308(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.lwoffi; PLASMID=pKLIH3, pKLIH102, AND pKLIH104;  
RA Kholodil G.Y., Mindlin S.Z., Lomovskaya O.L., Gorlenko Z.M.,  
RA Yurleva O.V., Nikiforov V.G.;

RT "pKLIH-like aberrant mercury resistance transposons of environmental  
RT Acinetobacter strains: spread, polymorphism and possible origin.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Acinetobacter sp.; STRAIN=ED45-25; PLASMID=pKLIH205;  
RX MEDLINE-21272500; PubMed-11376944;  
RA Kholodil G.Y.;  
RT "The shuffling function of resolvases.";  
RL Gene 269:121-130(2001).  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL: AF213017; AAA19678.1; -;  
DR EMBL: AJ251539; CAB65853.1; -;  
DR EMBL: AJ251517; CAB65939.1; -;  
DR EMBL: AJ250009; CAB65945.1; -;  
DR EMBL: AJ251537; CAB65949.1; -;  
DR EMBL: AJ251706; CAC39408.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr.1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR.1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
KW DNA-binding; Plasmid; Transcription regulation.  
SQ SEQUENCE 151 AA; 16561 MW; 549350FCE50240FC CRC64;  
Query Match 17.1%; Score 20; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 2, 9e-13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 LAEHLKDVREKMDLAME 29  
DB 88 LAEHLKDVREKMDLAME 107  
RESULT 15  
ID Q935L5 PRELIMINARY; PRT; 151 AA.  
AC Q935L5;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Putative mercury resistance operon regulatory protein.  
GN MERR.  
OS Salmonella typhi.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_Taxid=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE-21534947; PubMed-11677608;  
RA Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Baker S., Basham D., Brooks K., Bentley S.D., Holden M.T.G., Sebaila M.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL: AL513383; CAD09746.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr.1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY; UNKNOWN.1.  
KW DNA-binding; Plasmid; Transcription regulation; Complete proteome.  
SQ SEQUENCE 151 AA; 16589 MW; 239344ADA01754AD CRC64;

Query Match 17.1%; Score 20; DB 16; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-13;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LAEKKLDYREKMDLARME 29  
 ||||||||||||||||  
 Db 88 LAEKKLDYREKMDLARME 107

Search completed: May 28, 2003, 10:41:38  
 Job time : 85 secs



Mon Jun 2 10:57:32 2003

us-09-977-137a-4.oli.rai

Page 1

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:38:17 ; Search time 29 Seconds  
(without alignments)  
118.706 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 117  
Sequence: 1 MHCEASSIAEHKIKLVRE.....HARKGVSCPSAMSHPOFEK 117

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUTS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	6.8	8	3 US-08-948-097-2	Sequence 2, Appl
2	8	6.8	8	4 US-09-382-950-7	Sequence 7, Appl
3	8	6.8	8	4 US-09-382-736B-8	Sequence 8, Appl
4	8	6.8	8	4 US-09-619-103-9	Sequence 9, Appl
5	6	5.1	19	3 US-09-100-414B-105	Sequence 105, App
6	6	5.1	19	4 US-09-303-323-105	Sequence 105, App
7	6	5.1	25	6 5198345-12	Patent No. 5198345
8	6	5.1	31	3 US-09-100-414B-106	Sequence 106, App
9	6	5.1	31	4 US-09-303-323-106	Sequence 106, App
10	6	5.1	61	4 US-09-134-001C-4258	Sequence 4258, Ap
11	6	5.1	135	1 US-08-468-853-4	Sequence 4, Appl
12	6	5.1	135	1 US-08-468-855-4	Sequence 4, Appl
13	6	5.1	135	1 US-08-310-357-4	Sequence 4, Appl
14	6	5.1	135	1 US-08-468-852-4	Sequence 4, Appl
15	6	5.1	135	1 US-08-468-857-4	Sequence 18, Appl
16	6	5.1	187	3 US-08-737-248-18	Sequence 14, Appl
17	6	5.1	197	4 US-09-323-872A-14	Sequence 16, Appl
18	6	5.1	220	4 US-09-172-952-16	Sequence 28, Appl
19	6	5.1	220	4 US-09-172-952-28	Sequence 69, Appl
20	6	5.1	223	4 US-09-574-141A-69	Sequence 4606, Ap
21	6	5.1	246	4 US-09-134-001C-4606	Sequence 19, Appl
22	6	5.1	344	2 US-08-475-634D-19	Sequence 2, Appl
23	6	5.1	345	1 US-08-403-866-2	Sequence 8, Appl
24	6	5.1	350	2 US-08-960-022-8	Sequence 2, Appl
25	6	5.1	373	3 US-08-945-056-2	Sequence 5, Appl
26	6	5.1	590	2 US-08-785-310A-5	Sequence 2, Appl
27	6	5.1	640	3 US-09-026-343-2	Sequence 2, Appl

28	6	5.1	640	4 US-09-362-871-2	Sequence 2, Appl
29	6	5.1	983	2 US-08-164-292B-26	Sequence 26, Appl
30	6	5.1	983	3 US-08-845-623-26	Sequence 26, Appl
31	6	5.1	983	3 US-08-815-927-26	Sequence 26, Appl
32	6	5.1	983	4 US-09-103-330-26	Sequence 26, Appl
33	6	5.1	983	4 US-09-435-242-26	Sequence 26, Appl
34	6	5.1	1004	4 US-08-916-352-2	Sequence 4008, Ap
35	6	5.1	1162	4 US-09-134-001C-4008	Sequence 36, Appl
36	6	5.1	1247	3 US-08-058-489-35	Sequence 36, Appl
37	6	5.1	2547	3 US-09-058-489-36	Sequence 22, Appl
38	5	4.3	2555	3 US-08-918-148-22	Sequence 9, Appl
39	5	4.3	8	4 US-09-382-950-9	Sequence 10, Appl
40	5	4.3	8	4 US-09-382-736B-10	Sequence 25, Appl
41	5	4.3	9	4 US-09-053-941-25	Sequence 25, Appl
42	5	4.3	9	4 US-09-817-413-25	Sequence 78, Appl
43	5	4.3	13	4 US-08-981-392-78	Sequence 26, Appl
44	5	4.3	13	4 US-09-063-733A-26	Sequence 11, Appl
45	5	4.3	15	3 US-09-100-414B-11	

#### ALIGNMENTS

RESULT 1  
US-08-948-097-2  
; Sequence 2, Application US/08948097C  
; Patent No. 6103493  
; GENERAL INFORMATION:  
; APPLICANT: Skerter, Arne  
; APPLICANT: Voss, Selma  
; TITLE OF INVENTION: Streptavidin Mutelins  
; FILE REFERENCE: HUBR 1119  
; CURRENT APPLICATION NUMBER: US/08/948,097C  
; CURRENT FILING DATE: 1997-10-09  
; EARLIER APPLICATION NUMBER: DE 196 41 876.3  
; EARLIER FILING DATE: 1996-10-10  
; NUMBER OF SEQ ID NOS: 17  
; SEQ ID NO 2  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: BINDING  
; OTHER INFORMATION: Binding ligand for streptavidin  
US-08-948-097-2

Query Match 6.8% Score 8; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
|||||||  
Db 1 WSHPOFEK 8

RESULT 2  
US-09-382-950-7  
; Sequence 7, Application US/09382950  
; Patent No. 6303337  
; GENERAL INFORMATION:  
; APPLICANT: Rothschild, Kenneth  
; APPLICANT: Gile, Sadanand  
; APPLICANT: Olejnik, Jerzy  
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins  
; FILE REFERENCE: AMBER-03879  
; CURRENT APPLICATION NUMBER: US/09/382,950  
; CURRENT FILING DATE: 1999-08-25  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 7  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial/Unknown

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: ( )  
OTHER INFORMATION: Synthetic  
US-09-382-950-7

## Query Match

Best Local Similarity 6.8%; Score 8; DB 4; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

## RESULT 3

US-09-382-736B-8  
Sequence 8, Application US/09382736B  
Patent No. 6306628  
GENERAL INFORMATION:  
APPLICANT: Rothschild, Kenneth  
APPLICANT: Gite, Sadanand  
APPLICANT: Olejnik, Jerzy  
TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot  
FILE REFERENCE: AMBER-03951  
CURRENT APPLICATION NUMBER: US/09/382,736B  
CURRENT FILING DATE: 1999-08-25  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-382-736B-8

## Query Match

Best Local Similarity 6.8%; Score 8; DB 4; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

## RESULT 4

US-09-619-103-9  
Sequence 9, Application US/09619103  
Patent No. 6429300  
GENERAL INFORMATION:  
APPLICANT: Kurtz, Markus  
APPLICANT: Lohse, Peter  
APPLICANT: Wagner, Richard  
TITLE OF INVENTION: Peptide Acceptor Ligation Methods  
FILE REFERENCE: 50036/031002  
CURRENT APPLICATION NUMBER: US/09/619,103  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/145,834  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: designed sequence to act as an identifying tag  
US-09-619-103-9

Query Match  
Best Local Similarity 6.8%; Score 8; DB 4; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

RESULT 5  
US-09-100-414B-105  
Sequence 105, Application US/09100414B  
Patent No. 6025468

GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: NOVEL LHRH PEPTIDE  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Morgan & Flinnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154-0054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC Windows  
SOFTWARE: Word 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,414B  
FILING DATE: 20-JUNE-1998  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4157  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 105:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-100-414B-105

Query Match  
Best Local Similarity 5.1%; Score 6; DB 3; Length 19;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 RMEYVL 90  
Db 13 RMEYVL 18

## RESULT 6

US-09-303-323-105  
Sequence 105, Application US/09303323  
Patent No. 6228987  
GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: NOVEL LHRH PEPTIDE  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Morgan & Flinnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154-0054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC Windows  
SOFTWARE: Word 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/303,323  
FILING DATE: 30-APR-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/100,414  
FILING DATE: 20-JUNE-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4157  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 105:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-303-323-105

Query Match 5.1%; Score 6; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 RMEYVL 90  
|||||  
DB 13 RMEYVL 18

RESULT 7  
5198345-12  
PATENT NO. 5198345  
APPLICANT: GWYNNE, DAVID I.; BUXTON, FRANCIS P.; PICKETT, MARK H.  
DAVIES, ROGER W.; SCAZZOCHIO, CLAUDIO  
TITLE OF INVENTION: VECTORS IN USE IN FILAMENTOUS FUNGI  
NUMBER OF SEQUENCES: 28  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/811,404  
FILING DATE: 20-DEC-1985  
SEQ ID NO: 12  
LENGTH: 25  
5198345-12

Query Match 5.1%; Score 6; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 LSELVC 95  
|||||  
DB 9 LSELVC 14

RESULT 8  
US-09-100-414B-106  
Sequence 106, Application US/09100414B  
Patent No. 6025468  
GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: NOVEL LHRH PEPTIDE  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan & Finnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154-0054

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC Windows  
SOFTWARE: Word 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,414B  
FILING DATE: 20-JUNE-1998  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4157  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-100-414B-106

Query Match 5.1%; Score 6; DB 3; Length 31;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 RMEYVL 90  
|||||  
DB 13 RMEYVL 18

RESULT 9  
US-09-303-323-106  
Sequence 106, Application US/09303323  
Patent No. 6228987  
GENERAL INFORMATION:  
APPLICANT: Wang, Chang Yi  
TITLE OF INVENTION: NOVEL LHRH PEPTIDE  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morgan & Finnegan, L.L.P.  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10154-0054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC Windows  
SOFTWARE: Word 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/303,323  
FILING DATE: 30-APR-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/100,414  
FILING DATE: 20-JUNE-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4157  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-758-4800  
TELEFAX: 212-751-6849  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-09-303-323-106

Query Match  
Best Local Similarity: 100.0%; Score 6; DB 4; Length 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 RMETVL 90  
DB 13 RMETVL 18

RESULT 10  
US-09-134-001C-4258

Sequence 4258, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4258  
LENGTH: 61  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4258

Query Match  
Best Local Similarity: 100.0%; Score 6; DB 4; Length 61;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 PLIASL 54  
DB 50 PLIASL 55

RESULT 11  
US-08-468-853-4

Sequence 4, Application US/08468853  
Patent No. 5670362  
GENERAL INFORMATION:  
APPLICANT: van den BOOGAART, Paul  
APPLICANT: KOK, Jacobus Johannes  
APPLICANT: VERMEULEN, Arnoldus Nicolaas  
TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 5670362el Patent Department  
STREET: 1300 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,853  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,357  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/102,865  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,075  
FILING DATE: 18-JUN-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91.201.523.7  
FILING DATE: 18-JUN-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley

REGISTRATION NUMBER: 34,409  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-468-853-4

Query Match  
Best Local Similarity: 100.0%; Score 6; DB 1; Length 135;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EPASL 10  
DB 41 EPASL 46

RESULT 12  
US-08-468-855-4

Sequence 4, Application US/08468855  
Patent No. 5780289  
GENERAL INFORMATION:  
APPLICANT: van den BOOGAART, Paul  
APPLICANT: KOK, Jacobus Johannes  
APPLICANT: VERMEULEN, Arnoldus Nicolaas  
TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 5780289el Patent Department  
STREET: 1300 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,855  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,357  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/102,865  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,075  
FILING DATE: 18-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91.201.523.7  
FILING DATE: 18-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley  
REGISTRATION NUMBER: 34,409  
REFERENCE/DOCKET NUMBER:



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-468-855-4

Query Match 5.1%; Score 6; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EASL 10  
|||||  
DB 41 EASL 46

RESULT 13  
US-08-310-357-4  
Sequence 4, Application US/08310357  
Patent No. 5789233  
GENERAL INFORMATION:  
APPLICANT: van den BOOGAART, Paul  
APPLICANT: KOK, Jacobus Johannes  
TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Organon Teknika Corporation  
ADDRESSEE: Biotechnology Research Institute  
STREET: 1330-A Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/310,357  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/904,075  
FILING DATE:  
APPLICATION NUMBER: EP 91.201.523.7  
FILING DATE: 18-Jun-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: William M. Blackstone  
REGISTRATION NUMBER: 29,772  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-310-357-4

Query Match 5.1%; Score 6; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EASL 10  
|||||  
DB 41 EASL 46

RESULT 14  
US-08-468-852-4  
Sequence 4, Application US/08468852  
Patent No. 5792644  
GENERAL INFORMATION:  
APPLICANT: van den BOOGAART, Paul  
APPLICANT: KOK, Jacobus Johannes  
TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 5792644e1 Patent Department  
STREET: 1300 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,852  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,357  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/102,865  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,075  
FILING DATE: 18-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91.201.523.7  
FILING DATE: 18-Jun-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Maty E. Gormley  
REGISTRATION NUMBER: 34,409  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-468-852-4

Query Match 5.1%; Score 6; DB 1; Length 135;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EASL 10  
|||||  
DB 41 EASL 46

RESULT 15  
US-08-468-857-4  
Sequence 4, Application US/08468857  
Patent No. 5925347  
GENERAL INFORMATION:  
APPLICANT: van den BOOGAART, Paul  
APPLICANT: KOK, Jacobus Johannes  
TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Akzo No. 5925347e1 Patent Department  
STREET: 1300 Piccard Drive  
CITY: Rockville  
STATE: Maryland  
COUNTRY: U.S.A.  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,857  
FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/310,357  
FILING DATE: 21-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/102,865  
FILING DATE: 06-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/904,075  
FILING DATE: 18-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 91.201.523.7  
FILING DATE: 18-JUN-1991

ATTORNEY/AGENT INFORMATION:

NAME: Mary E. Gormley

REGISTRATION NUMBER: 34,409

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 135 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-468-857-4

Query Match 5.1%; Score 6; DB 2; Length 135;

Best Local Similarity 100.0%; Pred.No. 65;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EASL 10

DB 41 EASL 46

Search completed: May 28, 2003, 10:43:05  
Job time : 30 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 28, 2003, 10:41:43 ; Search time 57 Seconds  
(Without alignments)  
203.565 Million cell updates/sec

Title: US-09-977-137A-4

Sequence: 1 MTHCEASSLAHEHKLKDVRE.....HARKGVSCPSAMSHPOFEK 117

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 375593 seqs, 99172665 residues

Word size : 0

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCr\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/PCrUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	8.5	10	US-09-809-517A-6	Sequence 6, Appl1
2	8	6.8	8	US-10-104-218-5	Sequence 5, Appl1
3	8	6.8	8	US-10-208-357-9	Sequence 9, Appl1
4	8	6.8	8	US-10-001-934-8	Sequence 8, Appl1
5	8	6.8	8	US-10-026-578B-2	Sequence 2, Appl1
6	8	6.8	8	US-10-026-578B-9	Sequence 9, Appl1
7	8	6.8	8	US-10-174-368A-7	Sequence 7, Appl1
8	8	6.8	8	US-09-809-517A-9	Sequence 9, Appl1
9	8	6.8	8	US-09-973-145-7	Sequence 7, Appl1
10	8	6.8	9	US-09-983-067-3	Sequence 3, Appl1
11	8	6.8	21	US-09-809-517A-30	Sequence 30, Appl1
12	8	6.8	22	US-09-809-517A-33	Sequence 33, Appl1
13	8	6.8	24	US-10-026-578B-3	Sequence 3, Appl1
14	8	6.8	24	US-10-026-578B-4	Sequence 4, Appl1
15	8	6.8	24	US-09-809-517A-31	Sequence 31, Appl1
16	8	6.8	25	US-09-809-517A-34	Sequence 34, Appl1
17	8	6.8	36	US-10-026-578B-10	Sequence 10, Appl1
18	8	6.8	36	US-10-026-578B-11	Sequence 11, Appl1
19	7	6.0	161	US-10-103-313-359	Sequence 359, Appl1

20	7	6.0	192	9	US-09-764-891-5030	Sequence 5030, Ap
21	7	6.0	192	9	US-10-103-313-503	Sequence 503, App
22	7	6.0	500	9	US-10-078-770-138	Sequence 138, App
23	6	5.1	19	9	US-09-747-802-52	Sequence 52, Appl
24	6	5.1	19	9	US-09-865-294-44	Sequence 44, Appl
25	6	5.1	52	10	US-09-764-877-1962	Sequence 1962, Ap
26	6	5.1	54	9	US-09-989-920-192	Sequence 192, App
27	6	5.1	65	10	US-09-840-298-10	Sequence 10, Appl
28	6	5.1	65	10	US-09-840-298-11	Sequence 11, Appl
29	6	5.1	65	10	US-09-840-298-12	Sequence 12, Appl
30	6	5.1	70	10	US-09-764-877-1147	Sequence 1147, Ap
31	6	5.1	88	10	US-09-867-550-156	Sequence 156, App
32	6	5.1	93	10	US-10-091-504-1068	Sequence 1068, Ap
33	6	5.1	113	9	US-09-925-299-1271	Sequence 1271, Ap
34	6	5.1	113	10	US-09-925-299-1271	Sequence 1271, Ap
35	6	5.1	115	9	US-09-764-891-5242	Sequence 5242, Ap
36	6	5.1	115	9	US-09-764-891-5242	Sequence 13033, A
37	6	5.1	132	10	US-09-815-242-13033	Sequence 1018, Ap
38	6	5.1	136	1	US-08-976-063C-36	Sequence 36, Appl
39	6	5.1	159	10	US-09-764-864-1018	Sequence 1018, Ap
40	6	5.1	187	9	US-10-140-293-36	Sequence 36, Appl
41	6	5.1	187	9	US-10-140-293-37	Sequence 37, Appl
42	6	5.1	220	9	US-09-557-796-16	Sequence 16, Appl
43	6	5.1	220	9	US-09-557-796-28	Sequence 28, Appl
44	6	5.1	236	10	US-09-925-300-1191	Sequence 1191, Ap
45	6	5.1	260	9	US-09-764-868-992	Sequence 992, App

## ALIGNMENTS

RESULT 1  
US-09-809-517A-6  
; Sequence 6, Application US/09809517A  
; Patent No. US20020034733A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Corina  
; TITLE OF INVENTION: No. US20020034733A1e1 methods for displaying (poly)peptides/pr  
; TITLE OF INVENTION: Particles via disulfide bonds  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809, 517A  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-6  
Query Match 8.5%; Score 10; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00075;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 108 SAMSHPOFEK 117  
Db 1 SAMSHPOFEK 10  
RESULT 2  
US-10-104-218-5  
; Sequence 5, Application US/10104218  
; Patent No. US2002017196A1  
; GENERAL INFORMATION:  
; APPLICANT: MAIER, Thomas  
; APPLICANT: GABBERT, Carsten  
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS

FILE REFERENCE: MAIER, T. ET AL.-2  
CURRENT APPLICATION NUMBER: US/10/104,218  
CURRENT FILING DATE: 2002-03-22  
PRIORITY APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0  
PRIORITY FILING DATE: 2001-04-04  
PRIORITY APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0  
PRIORITY FILING DATE: 2001-05-03  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 5  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (1)..(8)  
OTHER INFORMATION: Streptagiti affinity peptide for protein purification  
US-10-104-218-5

Query Match  
Best Local Similarity 100.0%; Score 8; DB 9; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

RESULT 3  
US-10-208-357-9  
Sequence 9, Application US/10208357  
Publication No. US20020182687A1  
GENERAL INFORMATION:  
APPLICANT: Kurtz, Markus  
APPLICANT: Lohse, Peter  
TITLE OF INVENTION: Peptide Acceptor Ligation Methods  
FILE REFERENCE: 50036/031002  
CURRENT APPLICATION NUMBER: US/10/208,357  
CURRENT FILING DATE: 2002-07-30  
PRIORITY APPLICATION NUMBER: US/09/619,103  
PRIORITY FILING DATE: 2000-07-19  
PRIORITY APPLICATION NUMBER: 60/145,834  
PRIORITY FILING DATE: 1999-07-27  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: designed sequence to act as an identifying tag  
US-10-208-357-9

Query Match  
Best Local Similarity 100.0%; Score 8; DB 9; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

RESULT 4  
US-10-001-934-8  
Sequence 8, Application US/10001934  
Publication No. US20030032782A1  
GENERAL INFORMATION:  
APPLICANT: Nag, Zoltan  
TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE  
FILE OF INVENTION: KILLING OF CELLS INCLUDING LYMPHOID TUMOR CELLS  
FILE REFERENCE: GPGC-P01-003  
CURRENT APPLICATION NUMBER: US/10/001,934

CURRENT FILING DATE: 2001-11-15  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-10-001-934-8

Query Match  
Best Local Similarity 100.0%; Score 8; DB 9; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

RESULT 5  
US-10-026-578B-2  
Sequence 2, Application US/10026578B  
Publication No. US20030083474A1  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
APPLICANT: Schmidt, Thomas  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIORITY APPLICATION NUMBER: DE 101 13 776.1  
PRIORITY FILING DATE: 2001-03-21  
PRIORITY APPLICATION NUMBER: PCT/EP01/11846  
PRIORITY FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 8  
TYPE: PRT  
ORGANISM: artificial sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Peptide  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Artificial Sequence represents peptide binding module  
US-10-026-578B-2

Query Match  
Best Local Similarity 100.0%; Score 8; DB 9; Length 8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117  
Db 1 WSHPOFEK 8

RESULT 6  
US-10-026-578B-9  
Sequence 9, Application US/10026578B  
Publication No. US20030083474A1  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
APPLICANT: Schmidt, Thomas  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIORITY APPLICATION NUMBER: DE 101 13 776.1  
PRIORITY FILING DATE: 2001-03-21  
PRIORITY APPLICATION NUMBER: PCT/EP01/11846  
PRIORITY FILING DATE: 2001-10-12

```
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc_feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9
```

```
Query Match
Best Local Similarity 100.0%; Score 8; DB 9; Length 8;
Pred. No. 3.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 110 WSHPOFEK 117
|11111111|
DB 1 WSHPOFEK 8
```

```
RESULT 7
US-10-174-368A-7
; Sequence 7, Application US/10174368A
; Publication No. US20030092031A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
; FILE REFERENCE: AMBER-07145
; CURRENT APPLICATION NUMBER: US/10/174,368A
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/223233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc_feature
; OTHER INFORMATION: Synthetic
US-10-174-368A-7

Query Match
Best Local Similarity 100.0%; Score 8; DB 9; Length 8;
Pred. No. 3.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 WSHPOFEK 117
|11111111|
DB 1 WSHPOFEK 8

RESULT 8
US-09-809-517A-9
; Sequence 9, Application US/09809517A
; Patent No. US2002003473A1
; GENERAL INFORMATION:
; APPLICANT: Iohning, Corinna
; TITLE OF INVENTION: No. US2002003473A1el methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
```

```
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9
```

```
Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 8;
Pred. No. 3.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 110 WSHPOFEK 117
|11111111|
DB 1 WSHPOFEK 8
```

```
RESULT 9
US-09-973-145-7
; Sequence 7, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-06819
; CURRENT APPLICATION NUMBER: US/09/973,145
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 09/382,950
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc_feature
; OTHER INFORMATION: Synthetic
US-09-973-145-7
```

```
Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 8;
Pred. No. 3.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 110 WSHPOFEK 117
|11111111|
DB 1 WSHPOFEK 8
```

```
RESULT 10
US-09-983-067-3
; Sequence 3, Application US/09983067
; Patent No. US20020123101A1
; GENERAL INFORMATION:
; APPLICANT: INOUE, Akio
; APPLICANT: SHIMIZU, Yoshihiro
; APPLICANT: UEDA, Takuya
; TITLE OF INVENTION: Process for Producing Peptides By Using In Vitro
; FILE REFERENCE: 1752-0151P
; CURRENT APPLICATION NUMBER: US/09/983,067
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: JP 294795/2001
```

```
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: JP 227094/2001
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: JP 6910/2001
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 401417/2000
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strept-tag II used in the binding of streptavidin
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)-(9)
; OTHER INFORMATION: Schmidt et al., 1996, "Molecular interaction between the
; OTHER INFORMATION: Strept-tag affinity peptide and its cognate target,
; OTHER INFORMATION: streptavidin", J. of Mol. Biol. 255(5):753-766.
US-09-983-067-3
```

```
Query Match          6.8%; Score 8; DB 10; Length 9;
Best Local Similarity 100.0%; Pred.No.3.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 110 WSHPOFEK 117
      |||||
Db 2 WSHPOFEK 9
```

```
RESULT 11
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
```

```
Query Match          6.8%; Score 8; DB 10; Length 21;
Best Local Similarity 100.0%; Pred.No.0.17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 110 WSHPOFEK 117
      |||||
Db 14 WSHPOFEK 21
```

```
RESULT 12
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; OTHER INFORMATION: particles via disulfide bonds
```

```
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33
```

```
Query Match          6.8%; Score 8; DB 10; Length 22;
Best Local Similarity 100.0%; Pred.No.0.18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 110 WSHPOFEK 117
      |||||
Db 15 WSHPOFEK 22
```

```
RESULT 13
US-10-026-578B-3
; Sequence 3, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (10)..(10)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (11)..(11)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (12)..(12)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (13)..(13)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (14)..(14)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
```



Mon Jun 2 10:57:32 2003

us-09-977-137a-4.oli.rapb

Page 6

Job time : 58 secs

---



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:39:43 ; Search time 332 Seconds  
(Without alignments)  
227.210 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 117

Sequence: 1 MTHCEASSLAHKLKDYRE.....HARKGNVSCPSAMSHQFEK 117

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 4569144 seqs, 644733110 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	117	100.0	117	23	US-09-977-137A-4
2	64	54.7	117	23	US-09-977-137A-12
3	57	48.7	117	23	US-09-977-137A-5
4	57	48.7	117	23	US-09-977-137A-7
5	57	48.7	117	23	US-09-977-137A-8
6	57	48.7	117	23	US-09-977-137A-9

7	57	48.7	117	23	US-09-977-137A-10	Sequence 10, Appl
8	57	48.7 <td>117</td> <td>23<td>US-09-977-137A-11<td>Sequence 11, Appl</td></td></td>	117	23 <td>US-09-977-137A-11<td>Sequence 11, Appl</td></td>	US-09-977-137A-11 <td>Sequence 11, Appl</td>	Sequence 11, Appl
9	47.0	144	23	US-09-977-137A-2 <td>Sequence 2, Appl</td>	Sequence 2, Appl	
10	55	47.0	220	16	US-09-252-691-5823 <td>Sequence 5823, Ap</td>	Sequence 5823, Ap
11	55	47.0	220	16	US-09-252-691C-5823 <td>Sequence 5823, Ap</td>	Sequence 5823, Ap
12	55	47.0	221	18	US-09-489-039A-11501 <td>Sequence 11501, A</td>	Sequence 11501, A
13	55	47.0	255	16	US-09-252-691-6093 <td>Sequence 6093, Ap</td>	Sequence 6093, Ap
14	55	47.0	255	16	US-09-252-691C-6093 <td>Sequence 6093, Ap</td>	Sequence 6093, Ap
15	48	41.0	118	23	US-09-977-137A-6 <td>Sequence 6, Appl</td>	Sequence 6, Appl
16	22	18.8	121	16	US-09-252-991A-23151 <td>Sequence 23151, A</td>	Sequence 23151, A
17	22	18.8	151	16	US-09-252-991A-23125 <td>Sequence 23125, A</td>	Sequence 23125, A
18	22	18.8	151	16	US-09-252-991A-25251 <td>Sequence 25251, A</td>	Sequence 25251, A
19	19	8.5	10	22	US-09-809-517A-6 <td>Sequence 6, Appl</td>	Sequence 6, Appl
20	20	8	8	22	US-09-809-517A-9 <td>Sequence 9, Appl</td>	Sequence 9, Appl
21	8	6.8	8	23	US-09-973-145-7 <td>Sequence 7, Appl</td>	Sequence 7, Appl
22	8	6.8	8	24	US-10-001-934-8 <td>Sequence 8, Appl</td>	Sequence 8, Appl
23	8	6.8	8	24	US-10-049-332-8 <td>Sequence 8, Appl</td>	Sequence 8, Appl
24	8	6.8	8	25	US-10-104-218-5 <td>Sequence 5, Appl</td>	Sequence 5, Appl
25	8	6.8	8	26	US-10-208-357-9 <td>Sequence 9, Appl</td>	Sequence 9, Appl
26	8	6.8	9	23	US-09-983-067-3 <td>Sequence 3, Appl</td>	Sequence 3, Appl
27	8	6.8	21	22	US-09-809-517A-30 <td>Sequence 30, Appl</td>	Sequence 30, Appl
28	8	6.8	22	22	US-09-809-517A-33 <td>Sequence 33, Appl</td>	Sequence 33, Appl
29	8	6.8	24	22	US-09-809-517A-31 <td>Sequence 31, Appl</td>	Sequence 31, Appl
30	8	6.8	25	22	US-09-809-517A-34 <td>Sequence 34, Appl</td>	Sequence 34, Appl
31	8	6.8	32	23	US-09-979-284-19 <td>Sequence 19, Appl</td>	Sequence 19, Appl
32	8	6.8	205	23	US-09-980-862-24 <td>Sequence 24, Appl</td>	Sequence 24, Appl
33	8	6.8	254	23	US-09-980-862-20 <td>Sequence 20, Appl</td>	Sequence 20, Appl
34	8	6.8	294	23	US-09-980-862-22 <td>Sequence 22, Appl</td>	Sequence 22, Appl
35	8	6.8	396	23	US-09-980-862-19 <td>Sequence 19, Appl</td>	Sequence 19, Appl
36	8	6.8	645	24	US-10-015-127-10079 <td>Sequence 10079, A</td>	Sequence 10079, A
37	8	6.8	659	23	US-09-980-862-26 <td>Sequence 26, Appl</td>	Sequence 26, Appl
38	8	6.8	659	23	US-09-980-862-27 <td>Sequence 27, Appl</td>	Sequence 27, Appl
39	7	6.0	103	26	US-10-219-999-61542 <td>Sequence 61542, A</td>	Sequence 61542, A
40	7	6.0	103	27	US-60-312-544-10064 <td>Sequence 10064, A</td>	Sequence 10064, A
41	7	6.0	136	27	US-60-360-038-20227 <td>Sequence 20227, A</td>	Sequence 20227, A
42	7	6.0	158	19	US-09-513-996A-15593 <td>Sequence 15593, A</td>	Sequence 15593, A
43	7	6.0	158	20	US-09-620-394B-3514 <td>Sequence 3514, Ap</td>	Sequence 3514, Ap
44	7	6.0	161	1	PCT-US01-01358-359 <td>Sequence 359, App</td>	Sequence 359, App
45	7	6.0	161	21	US-09-764-854-359 <td>Sequence 359, App</td>	Sequence 359, App

#### ALIGNMENTS

RESULT 1  
US-09-977-137A-4  
Sequence 4, Application US/0977137A  
GENERAL INFORMATION:  
APPLICANT: Summers, Anne O.  
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
TITLE OF INVENTION: Methods  
FILE REFERENCE: 79-00  
CURRENT APPLICATION NUMBER: US/09/977,137A  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,465  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 117  
TYPE: PRN  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-4

Query Match 100.0%; Score 117; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 3.3e-113;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTHCEASSLAHKLKDYREKMDLARMETVLSLVCACIARRGNVSCPLIASIQSSGT 60  
|||||

Db 1 MTCEASSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPLIASLOGSSGT 60  
QY 61 HCEASSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPSANSHPOFEK 117  
Db 61 HCEASSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPSANSHPOFEK 117

## RESULT 2

US-09-977-137A-12  
; Sequence 12, Application US/09977137A

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 12

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-12

Query Match 54.7%; Score 64; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5e-58;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 LOGSSGTCEEASSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPSANSHP 113  
Db 54 LOGSSGTCEEASSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPSANSHP 113

QY 114 QFEK 117  
Db 114 QFEK 117

## RESULT 3

US-09-977-137A-5

; Sequence 5, Application US/09977137A

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 5

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-5

Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.8e-51;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 SSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPLIASLOGSSGTCEE 64  
Db 8 SSLAEHKIKDYREKMDLARMETVLSLVACACHARKGNVSCPLIASLOGSSGTCEE 64

RESULT 4  
US-09-977-137A-7  
; Sequence 7, Application US/09977137A

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 7

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-7

Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.8e-51;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 MADLARMETVLSLVACACHARKGNVSCPLIASLOGSSGTCEEASSLAEHKIKDYRE 78  
Db 22 MADLARMETVLSLVACACHARKGNVSCPLIASLOGSSGTCEEASSLAEHKIKDYRE 78

## RESULT 5

US-09-977-137A-8

; Sequence 8, Application US/09977137A

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 8

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-8

Query Match 48.7%; Score 57; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.8e-51;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 MADLARMETVLSLVACACHARKGNVSCPLIASLOGSSGTCEEASSLAEHKIKDYRE 78  
Db 22 MADLARMETVLSLVACACHARKGNVSCPLIASLOGSSGTCEEASSLAEHKIKDYRE 78

RESULT 6  
US-09-977-137A-9  
; Sequence 9, Application US/09977137A

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

```
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match          48.7%; Score 57; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 9.8e-51;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 ETVLSLVCAHARKGVSCPLIASLOGSSGTHCEASSLAHKLKDYREKMDIAR 85
Db 29 ETVLSLVCAHARKGVSCPLIASLOGSSGTHCEASSLAHKLKDYREKMDIAR 85

RESULT 7
US-09-977-137A-10
; Sequence 10, Application US/09977137A
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match          48.7%; Score 57; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 9.8e-51;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 ETVLSLVCAHARKGVSCPLIASLOGSSGTHCEASSLAHKLKDYREKMDIAR 85
Db 29 ETVLSLVCAHARKGVSCPLIASLOGSSGTHCEASSLAHKLKDYREKMDIAR 85

RESULT 8
US-09-977-137A-11
; Sequence 11, Application US/09977137A
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 11
```

```
; LENGTH: 117
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match          48.7%; Score 57; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 9.8e-51;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 CPLIASLOGSSGTHCEASSLAHKLKDYREKMDIARMETVLSLVCAHARKGV 104
Db 48 CPLIASLOGSSGTHCEASSLAHKLKDYREKMDIARMETVLSLVCAHARKGV 104

RESULT 9
US-09-977-137A-2
; Sequence 2, Application US/09977137A
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 144
; TYPE: PRF
; ORGANISM: Shigella flexneri, Tn21 of Plasmid R100
US-09-977-137A-2

Query Match          47.0%; Score 55; DB 23; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4e-48;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THCEASSLAHKLKDYREKMDIARMETVLSLVCAHARKGVSCPLIASLOG 56
Db 80 THCEASSLAHKLKDYREKMDIARMETVLSLVCAHARKGVSCPLIASLOG 134

RESULT 10
US-09-252-691-5823
; Sequence 5823, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROACTE
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 5823
; LENGTH: 220
; TYPE: PRF
; ORGANISM: Enterobacter cloacae
US-09-252-691-5823

Query Match          47.0%; Score 55; DB 16; Length 220;
Best Local Similarity 100.0%; Pred. No. 2.1e-48;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THCEASSLAHKLKDYREKMDIARMETVLSLVCAHARKGVSCPLIASLOG 56
Db 156 THCEASSLAHKLKDYREKMDIARMETVLSLVCAHARKGVSCPLIASLOG 210

RESULT 11
```

US-09-252-691C-5823  
; Sequence 5823, Application US/09252691C  
; GENERAL INFORMATION:  
; APPLICANT: Keith G. Weinstein et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER  
; FILE REFERENCE: 107196.135  
; CURRENT APPLICATION NUMBER: US/09/252,691C  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,145  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: US 60/074,787  
; PRIOR FILING DATE: 1998-02-18  
; NUMBER OF SEQ ID NOS: 11326  
; SEQ ID NO 5823  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Enterobacter cloacae  
US-09-252-691C-5823

Query Match 47.0%; Score 55; DB 16; Length 220;  
Best Local Similarity 100.0%; Pred. No. 2.1e-48;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THCEASSLAHKLKDVREKMDLARMETVSELVCACHARGNVSCPLIASLOG 56  
DB 156 THCEASSLAHKLKDVREKMDLARMETVSELVCACHARGNVSCPLIASLOG 210

## RESULT 12

US-09-489-039A-11501  
; Sequence 11501, Application US/09489039A  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KIEBSTIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 11501  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-11501

Query Match 47.0%; Score 55; DB 18; Length 221;  
Best Local Similarity 100.0%; Pred. No. 2.2e-48;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THCEASSLAHKLKDVREKMDLARMETVSELVCACHARGNVSCPLIASLOG 56  
DB 157 THCEASSLAHKLKDVREKMDLARMETVSELVCACHARGNVSCPLIASLOG 211

## RESULT 13

US-09-252-691-6093  
; Sequence 6093, Application US/09252691B  
; GENERAL INFORMATION:  
; APPLICANT: Keith G. Weinstein et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER  
; FILE REFERENCE: 107196.135  
; CURRENT APPLICATION NUMBER: US/09/252,691B  
; CURRENT FILING DATE: 1999-02-18  
; NUMBER OF SEQ ID NOS: 11324  
; SEQ ID NO 6093  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Enterobacter cloacae  
; FEATURE:

; NAME/KEY: UNSURE  
; LOCATION: (45)  
; NAME/KEY: UNSURE  
; LOCATION: (46)  
; NAME/KEY: UNSURE  
; LOCATION: (47)  
; NAME/KEY: UNSURE  
; LOCATION: (48)  
; NAME/KEY: UNSURE  
; LOCATION: (49)  
; NAME/KEY: UNSURE  
; LOCATION: (50)  
; NAME/KEY: UNSURE  
; LOCATION: (51)  
; NAME/KEY: UNSURE  
; LOCATION: (52)  
; NAME/KEY: UNSURE  
; LOCATION: (53)  
; NAME/KEY: UNSURE  
; LOCATION: (54)  
; NAME/KEY: UNSURE  
; LOCATION: (55)  
; NAME/KEY: UNSURE  
; LOCATION: (56)  
; NAME/KEY: UNSURE  
; LOCATION: (57)  
; NAME/KEY: UNSURE  
; LOCATION: (58)  
US-09-252-691-6093

Query Match 47.0%; Score 55; DB 16; Length 255;  
Best Local Similarity 100.0%; Pred. No. 2.5e-48;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THCEASSLAHKLKDVREKMDLARMETVSELVCACHARGNVSCPLIASLOG 56  
DB 191 THCEASSLAHKLKDVREKMDLARMETVSELVCACHARGNVSCPLIASLOG 245

## RESULT 14

US-09-252-691C-6093  
; Sequence 6093, Application US/09252691C  
; GENERAL INFORMATION:  
; APPLICANT: Keith G. Weinstein et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER  
; FILE REFERENCE: 107196.135  
; CURRENT APPLICATION NUMBER: US/09/252,691C  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,145  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: US 60/074,787  
; PRIOR FILING DATE: 1998-02-18  
; NUMBER OF SEQ ID NOS: 11326  
; SEQ ID NO 6093  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Enterobacter cloacae  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (45)  
; NAME/KEY: UNSURE  
; LOCATION: (46)  
; NAME/KEY: UNSURE  
; LOCATION: (47)  
; NAME/KEY: UNSURE  
; LOCATION: (48)  
; NAME/KEY: UNSURE  
; LOCATION: (49)  
; NAME/KEY: UNSURE  
; LOCATION: (50)  
; NAME/KEY: UNSURE  
; LOCATION: (51)

NAME/KEY: UNSURE  
LOCATION: (52)  
NAME/KEY: UNSURE  
LOCATION: (53)  
NAME/KEY: UNSURE  
LOCATION: (54)  
NAME/KEY: UNSURE  
LOCATION: (55)  
NAME/KEY: UNSURE  
LOCATION: (56)  
NAME/KEY: UNSURE  
LOCATION: (57)  
NAME/KEY: UNSURE  
LOCATION: (58)  
US-09-252-691C-6093

Query Match  
Best Local Similarity 47.0%; Score 55; DB 16; Length 255;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

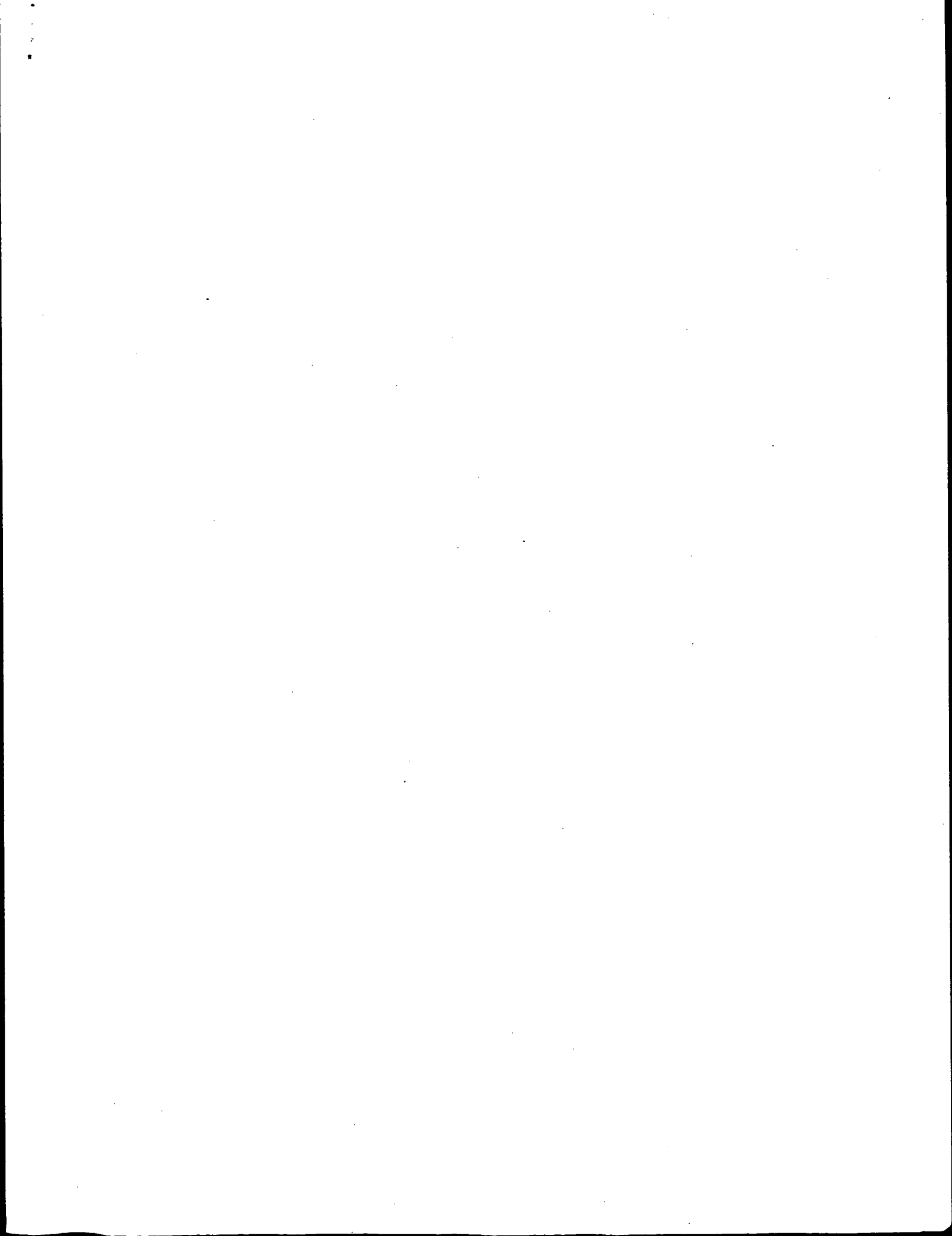
OY 2 THCEASLSAEHLKDYREKMDIARMEYVLSVLCACHARKGNVSCPLIASLOG 56  
DB 191 THCEASLSAEHLKDYREKMDIARMEYVLSVLCACHARKGNVSCPLIASLOG 245

RESULT 15  
US-09-977-137A-6  
; Sequence 6, Application US/09977137A  
; GENERAL INFORMATION:  
; APPLICANT: Summers, Anne O.  
; APPLICANT: Caguiat, Jonathan  
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
; FILE REFERENCE: 79-00  
; CURRENT APPLICATION NUMBER: US/09/977,137A  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,465  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 118  
; TYPE: PRP  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-6

Query Match  
Best Local Similarity 41.0%; Score 48; DB 23; Length 118;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 EHLKDYREKMDIARMEYVLSVLCACHARKGNVSCPSAMSHPOFEK 117  
DB 71 EHLKDYREKMDIARMEYVLSVLCACHARKGNVSCPSAMSHPOFEK 118

Search completed: May 28, 2003, 10:48:43  
Job time : 332 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 05:46:59 ; Search time 47 Seconds  
(without alignments)  
2094.537 Million cell updates/sec

Title: US-09-977-137A-3  
Perfect score: 321  
Sequence: 1 atgacacactgcgagagagc.....agggaattctctccgcgcg 321

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents\_NA :  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.2	11.3	2799	4 US-09-232-279-1	Sequence 1, Appli
2	36.2	11.3	3382	2 US-08-682-847-1	Sequence 1, Appli
3	36.2	11.3	3519	1 US-08-035-558-1	Sequence 1, Appli
4	35.4	11.0	289	4 US-09-007-005-17	Sequence 17, Appl
5	35.4	11.0	289	4 US-09-244-796-17	Sequence 17, Appl
6	34.2	10.7	4403765	4 US-09-103-840A-2	Sequence 2, Appli
7	34.2	10.7	4411529	4 US-09-103-840A-1	Sequence 1, Appli
8	33.8	10.5	20235	1 US-07-642-734C-3	Sequence 3, Appli
9	33.8	10.3	2277	1 US-08-439-009A-3	Sequence 3, Appli
10	33.2	10.3	2277	1 US-08-676-967-5	Sequence 5, Appli
11	33.2	10.3	2277	1 US-08-676-974-5	Sequence 5, Appli
12	33.2	10.3	2277	2 US-09-098-487-5	Sequence 5, Appli
13	33	10.3	2838	4 US-08-246-489-1	Sequence 1, Appli
14	32.8	10.2	1299	4 US-09-199-637A-160	Sequence 160, App
15	32.6	10.2	1777	2 US-08-173-508-5	Sequence 5, Appli
16	32.6	10.2	1777	2 US-08-265-310-5	Sequence 5, Appli
17	32.6	10.2	1777	2 US-08-951-742-5	Sequence 5, Appli
18	31.2	9.7	1341	4 US-08-199-637A-133	Sequence 133, App
19	31.2	9.7	2051	4 US-09-199-637A-114	Sequence 114, App
20	31.2	9.7	2757	5 US-08-306-691B-48	Sequence 48, Appl
21	31.2	9.7	15872	4 PCT-US93-06251-79	Sequence 79, Appl
22	31.2	9.7	15872	4 US-09-105-537-1	Sequence 1, Appli
23	30.8	9.6	4403765	4 US-09-103-840A-2	Sequence 2, Appli
24	30.8	9.6	4411529	4 US-09-103-840A-1	Sequence 1, Appli
25	30.6	9.5	516	4 US-09-889-595-4	Sequence 4, Appli
26	30.6	9.5	1155	4 US-08-818-112-12	Sequence 12, Appl
27	30.6	9.5	1155	4 US-08-818-111-12	Sequence 12, Appl

28	30.6	9.5	1155	4 US-09-056-556-12	Sequence 12, Appl
29	30.6	9.5	1155	4 US-09-072-596-12	Sequence 12, Appl
30	30.6	9.5	4171	4 US-09-754-250-1	Sequence 1, Appli
31	30.6	9.5	5058	4 US-09-889-595-1	Sequence 1, Appli
32	30.4	9.5	1151	2 US-08-807-044-2	Sequence 2, Appli
33	30.4	9.5	2943	1 US-08-042-747A-7	Sequence 7, Appli
34	30.4	9.5	3027	2 US-08-680-326-23	Sequence 23, Appl
35	30.4	9.5	23673	4 US-09-773-816-1	Sequence 1, Appli
36	30.2	9.4	3760	2 US-08-724-354D-1	Sequence 1, Appli
37	30.2	9.4	3760	3 US-09-370-984A-1	Sequence 1, Appli
38	29.8	9.3	1087	4 US-09-370-422A-29	Sequence 29, Appl
39	29.8	9.3	2505	1 US-07-977-434-7	Sequence 7, Appli
40	29.8	9.3	2505	1 US-08-458-819-7	Sequence 7, Appli
41	29.8	9.3	2505	5 PCT-US91-07035-7	Sequence 7, Appli
42	29.8	9.3	2713	3 US-08-804-439A-13	Sequence 13, Appl
43	29.8	9.3	2713	3 US-08-720-229-13	Sequence 13, Appl
44	29.8	9.3	3472	6 US-07-92-2	Patent No. 5244792
45	29.8	9.3	11236	1 US-07-853-913-1	Sequence 1, Appli

## ALIGNMENTS

```
RESULT 1
US-09-232-279-1
; Sequence 1, Application US/09232279
; Patent No. 6376473
; GENERAL INFORMATION:
; APPLICANT: Audonnet et al.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA, PARTICULARLY FOR
; TREATING BOVINE RESPIRATORY DISEASE
; FILE REFERENCE: 454313-2250
; CURRENT APPLICATION NUMBER: US/09/232,279
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2799
; TYPE: DNA
; ORGANISM: Bovine herpesvirus 2
US-09-232-279-1

Query Match          11.3%; Score 36.2; DB 4; Length 2799;
Best Local Similarity 47.6%; Pred. No. 0.28;
Matches 107; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY      9 CTGCGAGGAGGCCAGCCTGGCCGCAACAAGCTCAAGAGCTGCCGGAAGATGCC 68
      || || || || || || || || || || || || || || || || || || || ||
DB      1911 CTTCCGCAACGAGAGCGAGCGCGGTGGAGGGCCAGCTCGCGAGCAACGCTGCGCC 1970

QY      69 CCACTTGGCGCGCATGTGAACCGTGTGTAAGTGTGTGCGCCCATGCACGAAA 128
      || || || || || || || || || || || || || || || || || || || ||
DB      1971 GGGCCCGGAGCTGTGTGAGCCCTGCACCGCCAAACCAACGACTTCTTGGCGC 2030

QY      129 GGGGAATGTTTCCCTGCGCCGTGTGATCGGCTACTACAGGATCTCCAGGACCCACTGCGA 188
      || || || || || || || || || || || || || || || || || || || ||
DB      2031 GAGCTACGTGTATATACGAGAACTACGCGTACGCGGGGTCCCGTCCGCGGAGCTCGGA 2090

QY      189 GGAGCGCAGCAGCGCTGGCCGCAACAAGCTCAAGAGAGCTGCGCGCA 233
      || || || || || || || || || || || || || || || || || || || ||
DB      2091 GGTGATCAGCAGCCTTGTGTGACCTTAACCTACAGGTTCTGTGAGGA 2135

RESULT 2
US-08-682-847-1
; Sequence 1, Application US/08682847
; Patent No. 5858989
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE
; APPLICANT: VAN DEN HURK, SYLVIA
; APPLICANT: ZAMB, TIM
; APPLICANT: FITZPATRICK, DAVID
; TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1
```

```

1 TITLE OF INVENTION: POLYPEPTIDES AND VACCINES
2 NUMBER OF SEQUENCES: 8
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: MORRISON & FOERSTER
5 STREET: 755 PAGE MILL ROAD
6 CITY: PALO ALTO
7 STATE: CA
8 COUNTRY: USA
9 ZIP: 94304-1018
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/682,847
17 FILING DATE: 12-JUL-1996
18 CLASSIFICATION: 536
19 ATTORNEY/AGENT INFORMATION:
20 NAME: PARK, FREDDIE K.
21 REGISTRATION NUMBER: 35,636
22 REFERENCE/DOCKET NUMBER: 29310-20005.10
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (415) 813-5600
25 TELEFAX: (415) 494-0792
26 TELEX: 706141
27 INFORMATION FOR SEQ ID NO: 1:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 3382 base pairs
30 TYPE: nucleic acid
31 STRANDEDNESS: double
32 TOPOLOGY: linear
33 MOLECULE TYPE: DNA (genomic)
34 FEATURE:
35 NAME/KEY: CDS
36 LOCATION: 432..3230
37 US-08-682-847-1
38
39 Query Match 11.3%; Score 36.2; DB 2; Length 3382;
40 Best Local Similarity 47.6%; Pred. No. 0.25;
41 Matches 107; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
42
43 QY 9 CTGCGAGGAGGCGCCAGCAGCTCGGCCGAACACAGCTCAAGGACGTCGCGAGAGATGGC 68
44 DB 2345 CTTGGCCAGCAGAGCGAGCGCGGTGGAGGGCCAGCTCGGCCAGAGAACAGAGCTGTCGCC 2404
45 QY 69 CGACTTGGCGCGCGATGGAACCGTGCTGTCTGAAGTCTGTTGGCGCTTCGATCGACGAA 128
46 DB 2405 GGGCCCGAGCGTGTGTGAGGCCCTGCACCGCCAAACCAAGCCGCTTACCTTCGCTTTGGCGC 2464
47 QY 129 GGGGAAATGTTTCTCGCCCGTTGATCGCGTCACTACAGGAGATCCTTCAGGGACCCCACTGCG 188
48 DB 2465 GGACTACAGTGTACTACAGAACTACGCTAGCTGGCGGGGTGCCCTCGCGGAGCTGA 2524
49 QY 189 GGAGGCCAGCAGCCTGGCCGGAACAACAAGCTCAAGGACGTGCGCGA 233
50 DB 2525 GGTGATCAGCAGCAGCTTGTGTGAGACCTCAACCTCAGCTTGTGGAGGA 2569
51
52 RESULT 3
53 US-08-035-558-1
54 Sequence 1, Application US/08035558
55 Patent No. 5462734
56 GENERAL INFORMATION:
57 APPLICANT: Letchworth, Geoffrey J.
58 APPLICANT: Israel, Barbara A.
59 TITLE OF INVENTION: HERPESVIRUS VACCINE AND METHOD
60 TITLE OF INVENTION: OF USING SAME
61 NUMBER OF SEQUENCES: 3
62 CORRESPONDENCE ADDRESS:
63 ADDRESSEE: Quarles and Brady
64 STREET: P O BOX 2113
65 STREET: FIRST WISCONSIN PLAZA

```

```

1 CITY: MADISON
2 STATE: WISCONSIN
3 COUNTRY: U.S.A.
4 ZIP: 53701
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
7 COMPUTER: Apple Macintosh
8 OPERATING SYSTEM: Macintosh
9 SOFTWARE: Microsoft Word 4.0
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/035.558
12 FILING DATE:
13 CLASSIFICATION: 424
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: 07/607,794
16 FILING DATE: 02-NOV-1990
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Nicholas J. Seay
19 REGISTRATION NUMBER: 27,386
20 REFERENCE/DOCKET NUMBER: 9629691682
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (608)251-5000
23 TELEFAX: (608)251-9166
24 INFORMATION FOR SEQ ID NO: 1:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 3519 base pairs
27 TYPE: nucleic acid
28 STRANDEDNESS: double
29 TOPOLOGY: linear
30 MOLECULE TYPE: genomic DNA
31 HYPOTHEICAL: no
32 ANTI-SENSE: no
33 ORIGINAL SOURCE:
34 ORGANISM: bovine herpesvirus-1
35 INDIVIDUAL ISOLATE: P8-2 strain
36 FEATURE:
37 NAME/KEY: coding sequence for gI glycoprotein
38 LOCATION: 379 to 3165
39 IDENTIFICATION METHOD: sequence analysis
40 PUBLICATION INFORMATION:
41 AUTHORS: Mistra, Vikram
42 AUTHORS: Nelson, Randy
43 AUTHORS: Smith, Michael
44 TITLE: Sequence of a Bovine Herpesvirus Type-1 etc.
45 JOURNAL: Virology
46 VOLUME: 166
47 PAGES: 542-549
48 DATE: 1988
49 RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 379 TO 3165
50 US-08-035-558-1
51
52 Query Match 11.3%; Score 36.2; DB 1; Length 3519;
53 Best Local Similarity 47.6%; Pred. No. 0.3; Mismatches 118; Indels 0; Gaps 0;
54 Matches 107; Conservative 0;
55
56 9 CTGCGAGAGGCGCAGCAGCTGCGCGAACAACAAGCTCAAGGACGTGCGGAGAAAGATGGC 68
57 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
58 2280 CTTCGGCAAGAGAGAGCGGCGGTGGAGGGCGCAGCTCGCGGAGACAACAAGCTGTGCC 2339
59 CGACTTGCGCGCATGGAACCGTGTCTGAACTCGTGTGCGCGCTGCCATGCAACGAAA 128
60 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
61 2340 GGGCGGAGAGCTGTGGAGCCCTCACCAGCCAAACACAAAGGCTACTTCCGCTTGGCGC 2399
62 129 GGGGAAGTGTCCCGCCCGTGTGATCGGTGCTACTACAGGATCTTCAGGACACCACCTGCGA 188
63 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
64 2400 GGACTACGTACTACAGAGAACTACGCTACGTACGTGCGGGGTCCCGCTGCGGGAGCTGGA 2459
65 189 GGAGCGCAGCAGCTGCGCGAACAACAAGCTCAAGGACGTGCGGCA 233
66 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
67 2460 GGTATCATGACACTTTGTGTGACCTTAACCTACGCTTCTGTGAGGA 2504

```



```

US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patient No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Szostak, Richard W.
; APPLICANT: Liu, Rihé
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007, 005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035, 963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064, 491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc-feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 11.0%; Score 35.4; DB 4; Length 289;
Best Local Similarity 5.0%; Pred. No. 0.3;
Matches 11; Conservative 98; Mismatches 112; Indels 0; Gaps 0

QY 40 AACGCTCAAGCAGCGCGCGAAGAATGCGCACTTGGCGCCATGAACCGTGTCT 99
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 33 RARCRARARURURARCARARARURGRNRRNRRNRRNRRNRRNRRNRRNRRNRRN 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 100 GAACCTGCTGTCGCGCTCGCATGACAGAAAGGGAAATGTTTCTGCGCGTTCG 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 160 CTACAGGATCTCAGGACCCACGCGAGGAGGAGCCAGCAGCTGCGCCAGACAGCT 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 220 AAGGACGTGCGCGAAGAGATGCGCGACTTGGCGCGCATGA 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 RNRNRSRCRARGRCRURGRCRURBARARCRURCRURGR 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihé
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244, 796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035, 963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064, 491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007, 005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17

```

```

: LENGTH: 289
: TYPE: RNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(289)
: OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          11.0%; Score 35.4; DB 4; Length 289;
Best Local Similarity 5.0%; Pred. No. 0.3;
Matches 11; Conservative 98; Mismatches 112; Indels 0; Gaps 0.

QY  40  AAGCTCAAGACGTCGCCGAGACAGATGGCGCACTTGGCGGCATGGAACCGTCTCTCT 99
    :|: |: : : |: :|: |: : : : : : : : : : : : : : : : :
Db   33  RARCARARURURARCARARURGRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 92

QY  100 GAATCGTGTGCGCCTGCCCATGCACGAAAGGGGAATGTTCTCTGCCCTTATGCCGTC 159
    :|: |: : : : : : : : : : : : : : : : : : : : : :
Db   93  RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 152

QY  160 CTACGGAGATCTCTGCGCACCCACATCGCAGGAGGCGCACAGCTGCGCGAACACAGCTC 219
    :|: |: : : : : : : : : : : : : : : : : : : : : :
Db   153 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 212

QY  220 AAGCAGCTGCGAGAGATGCGCCACTTGGCGGCATGGA 260
    :|: |: : : : : |: :|: |: : : : : : : : : : : :
Db   213 RNRNRSRCRARCRURGRCRURURARARCRURCRURRGR 253

```

```

RESULT 6
US-09-103-840A-2
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: Patentln Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          10.7%; Score 34.2; DB 4; Length 4403765;
Best Local Similarity 48.2%; Pred. No. 4.1;
Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Df 16 GAGCGCCAGCAGCCTGGCCGACACACAGCTCAAGAGCGTGGCGGAGAGATGGCCGACTTG 75
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
791022 GTGGTGCGGTGGGGGATGGCCGCGCAGGTGATGAAAGTGTCGGCGGCAAGATCGCCGGGCG 791081

Df 76 GCGCGCATGGAACCGTGGCTGTCTGACTGCTGTGGCGCTGCGCATCAGCAAGGGGAAT 135
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
791082 GCGCGCATGCGCGAATGTGTAGCCCAAACTCAAGCGCTAAGAGCTGAGCTAGGTGGACGG 791141

Df 136 GTTTCCTGGCCGTGATTCGCGTACATACAGAGGATCTCAAGGCAACCACTGCGAGGAGCC 195
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
791142 GCGCCTGTGTCGTGCGCGCATTCGCGGAGCATATCTGCTGTGGGTGGGAGCGGCTGGCC 791201

```

QY 196 AGCAGCCTGGCCGACACA 214  
Db 791202 AGCGCGCGCGCGCACGCA 791220

## RESULT 7

US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, John C.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103, 840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 10.7%; Score 34.2; DB 4; Length 4411529;  
Best Local Similarity 48.2%; Pred. No. 4.1; Mismatches 103; Indels 0; Gaps 0;  
Matches 96; Conservative 0;

QY 16 GAGCCAGCAGCCTGGCCGACACAAGCTCAAGAGCAGTGGCGAGAGATGGCCGACTTG 75  
Db 789060 GTGTGCTGGTGGCGGATGGCCGAGTGTATGAGTGAAGTGGCGGCAATGGCGCGGC 789119  
76 GCGCGCATGGAACCGCTGTCTGACTGTGTGGCGCCCTGACCATGACCAAGGGGAT 135  
Db 789120 GCGCCCATGCGCGGAAGTGTAGACCAACTCAAGGCTTAGAGCTGACCTAGGTGGCAGCG 789179  
QY 136 GTTTCCTGCCCGCTGTGATGCGCTACTACAGAGGATCTCAAGCAGCAGCTCGAGAGGAGCC 195  
Db 789180 GCGCCCTGTGTGTGGCGCATGCGCGGACATATGCTGTGGGAGCAGCGCTCGGCC 789239  
QY 196 AGCAGCCTGGCCGACACA 214  
Db 789240 AGCGCGCGCGCGCACGCA 789258

## RESULT 8

US-07-642-734C-3  
Sequence 3, Application US/07642734C  
Patent No. 5824513  
GENERAL INFORMATION:  
APPLICANT: Katz, L.  
APPLICANT: Donadio, S.  
APPLICANT: McAlpine, J B  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
TITLE OF INVENTION: Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Edward H. Gorman  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckers, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRL 238  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..10722  
OTHER INFORMATION: /codon\_start= 19  
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 19..4470  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 97..1482  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1693..2670  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3406..3921  
OTHER INFORMATION: beta-ketoreductase domain of module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4171..4428  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 3"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4471..10722  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 4"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4471..5847  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6054..7026  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 4"  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION:	7165..9216	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		dehydratase and enoylreductase domains m"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	9433..9984	/function-	"approximate span
OTHER INFORMATION:			
OTHER INFORMATION:		beta-ketoreductase of module 4"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	10225..10483	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		acyl carrier domain of module 4"	
FEATURE:			
NAME/KEY:	CDS		
LOCATION:	10723..20235		
OTHER INFORMATION:	/codon_start= 10723		
OTHER INFORMATION:	/function="gene="eryA"		
OTHER INFORMATION:	/product="orf3 encoding modules 5 & 6		
OTHER INFORMATION:	6-deoxyerythronolide B formatio"		
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	10723..15165	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		module 5"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	10831..12174	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		beta-ketoacylACPSynthase domain of modul"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	12379..13350	/function-	"approximatr span of
OTHER INFORMATION:			
OTHER INFORMATION:		acylttransferase domain of module 5"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	14062..14610	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		beta-ketoreductase of module 5"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	14857..15114		
OTHER INFORMATION:		/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		acyl carrier domain of module 5"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	15166..20235	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		module 6"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	15172..16569	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		beta-ketoacylACPSynthase domain of modul"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	16768..17721	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		acylttransferase domain of module 6"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	18379..18921	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		beta-ketoreductase domain of module 6"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	19149..19398	/function-	"approximate span of
OTHER INFORMATION:			
OTHER INFORMATION:		acyl carrier domain of module 6"	
FEATURE:			
NAME/KEY:	misc_feature		
LOCATION:	19492..20235		

```

; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match      10.5%;   score 33.8;  DB 1;   Length 20235;
Best Local Similarity 53.0%;   Pred. No. 1.9;
Matches 96;  Conservative 0;  Mismatches 82;  Indels 3;  Gaps 1;

QY      21  CAGCAGCCTGGCCGACACACAAGCTCAAGAGCAGTGGCGCGAGAGATGGCGGACTTTGGCGCG 80
      |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db      3654 CGGACGCACTACGCGAAGTGGCAGCCGAGAGATTTGGCCGACAGACATCGGGCGCAAGACCCG 3713

QY      81  CATGGAACCGCTGCTGTCTGACTGTGTGGCGCTGCCATGCACGAAAGGGGAATGTTTC 140
      ||  |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db      3714 GTTGCTGCACCCTGGTCGACGAAGTCCCT---CGCGACCGGGGCGCTGCGAGCGGAGGTCTA 3770

QY      141 CTGCCCCTGTATTCGCGTCACTACAGAGATCTTCAGGCACCACTGCGAGAGGCCAGAG 200
      ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db      3771 CTGCTCTCGGTGGCCCGGGGATGTGGGGCGGGCGGGATGGCCGCTTACGCGGACAGCGAG 3830

QY      201 C 201
      |
Db      3831 C 3831

```

```

1      RESULT 9
2      US-08-439-009A-3
3      Sequence 3, Application US/08439009A
4      Patent No. 6004787
5      GENERAL INFORMATION:
6      APPLICANT: Donadio, S
7      APPLICANT: Katz, L
8      APPLICANT: McAlpine, J B
9      TITLE OF INVENTION: Method of Directing Biosynthesis of
10     TITLE OF INVENTION: Specific Polyketides
11     NUMBER OF SEQUENCES: 27
12     CORRESPONDENCE ADDRESS:
13     ADDRESSEE: Steven F. Weinstein
14     STREET: Abbott Laboratories D377/Ar6D-2 One Abbott
15     STREET: Park Rd
16     CITY: Abbott Park
17     STATE: IL
18     COUNTRY: US
19     ZIP: 60064-3500
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: Floppy disk
22     COMPUTER: IBM PC compatible
23     OPERATING SYSTEM: PC-DOS/MS-DOS
24     SOFTWARE: PatentIn Release #1.0, Version #1.25
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/08/439,009A
27     FILING DATE: 11-MAY-1995
28     CLASSIFICATION: 435
29     ATTORNEY/AGENT INFORMATION:
30     NAME: Casuto, Dianne
31     REGISTRATION NUMBER: 40,943
32     REFERENCE/DOCKET NUMBER: 4952.US.D1
33     TELECOMMUNICATION INFORMATION:
34     TELEPHONE: 847-938-3137
35     INFORMATION FOR SEQ ID NO: 3:
36     SEQUENCE CHARACTERISTICS:
37     LENGTH: 20235 base pairs
38     TYPE: nucleic acid
39     STRANDEDNESS: double
40     TOPOLOGY: unknown
41     MOLECULE TYPE: DNA (genomic)
42     HYPOTHETICAL: NO
43     ANTI-SENSE: NO
44     ORIGINAL SOURCE:
45     ORGANISM: Saccharopolyspora erythraea
46     STRAIN: NRRL 238
47     FEATURE:
48     NAME/KEY: CDS

```

```

1 LOCATION: 19..10722 /codon_start-19
2 OTHER INFORMATION: /function-"gene eryA"
3 OTHER INFORMATION: /product-"eryA ORF2 encoding modules 3 & 4 for
4 OTHER INFORMATION: 6-deoxyerythronolide B"
5 FEATURE:
6 NAME/KEY: misc_feature
7 LOCATION: 19..4470
8 OTHER INFORMATION: /function-"approximate span of
9 OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: 97..1482
13 OTHER INFORMATION: /function-"approximate span of
14 OTHER INFORMATION: acyltransferase domain module 3"
15 FEATURE:
16 NAME/KEY: misc_feature
17 LOCATION: 3406..3921
18 OTHER INFORMATION: /function-"approximate span of
19 OTHER INFORMATION: beta-ketoreductase domain of module 3"
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: 4171..4428
23 OTHER INFORMATION: /function-"approximate span of
24 OTHER INFORMATION: acyl carrier domain of module 3"
25 FEATURE:
26 NAME/KEY: misc_feature
27 LOCATION: 4471..10722
28 OTHER INFORMATION: /function-"approximate span of
29 OTHER INFORMATION: module 4"
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: 6054..7026
33 OTHER INFORMATION: /function-"approximate span of
34 OTHER INFORMATION: acyltransferase domain of module 4"
35 FEATURE:
36 NAME/KEY: misc_feature
37 LOCATION: 7165..9216
38 OTHER INFORMATION: /function-"approximate span of
39 OTHER INFORMATION: dehydratase and enoylreductase domains m"
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: 9453..9984
43 OTHER INFORMATION: /function-"approximate span
44 OTHER INFORMATION: beta-ketoreductase of module 4"
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: 10225..10483
48 OTHER INFORMATION: /function-"approximate span of
49 OTHER INFORMATION: acyl carrier domain of module 4"
50 FEATURE:
51 NAME/KEY: CDS
52 LOCATION: 10723..10235
53 OTHER INFORMATION: /codon_start-10723
54 OTHER INFORMATION: /function-"gene -eryA"
55 OTHER INFORMATION: /product-"orf3 encoding modules 5 & 6
56 OTHER INFORMATION: 6-deoxyerythronolide B formatio"
57 FEATURE:
58 NAME/KEY: misc_feature
59 LOCATION: 10723..15165
60 OTHER INFORMATION: /function-"approximate span of
61 OTHER INFORMATION: module 5"
62 FEATURE:

```

```

1 NAME/KEY: misc_feature
2 LOCATION: 10831..12174
3 OTHER INFORMATION: /function= "approximate span of
4 OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
5 FEATURE:
6 NAME/KEY: misc_feature
7 LOCATION: 12379..13350
8 OTHER INFORMATION: /function= "approximatr span of
9 OTHER INFORMATION: acyltransferase domain of module 5"
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: 14062..14610
13 OTHER INFORMATION: /function= "approximate span of
14 OTHER INFORMATION: beta-ketoreductase of module 5"
15 FEATURE:
16 NAME/KEY: misc_feature
17 LOCATION: 14857..15114
18 OTHER INFORMATION: /function= "approximate span of
19 OTHER INFORMATION: acyl carrier domain of module 5"
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: 15166..20235
23 OTHER INFORMATION: /function= "approximate span of
24 OTHER INFORMATION: module 6"
25 FEATURE:
26 NAME/KEY: misc_feature
27 LOCATION: 15172..16569
28 OTHER INFORMATION: /function= "approximate span of
29 OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: 16768..17721
33 OTHER INFORMATION: /function= "approximate span of
34 OTHER INFORMATION: acyltransferase domain of module 6"
35 FEATURE:
36 NAME/KEY: misc_feature
37 LOCATION: 18379..18921
38 OTHER INFORMATION: /function= "approximate span of
39 OTHER INFORMATION: beta-ketoreductase domain of module 6"
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: 19149..19398
43 OTHER INFORMATION: /function= "approximate span of
44 OTHER INFORMATION: acyl carrier domain of module 6"
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: 19492..20235
48 OTHER INFORMATION: /function= "approximate span of
49 OTHER INFORMATION: thioesterase domain of module 6"
50
51 US-08-439-009A-3

```

Query Match	Similarity	10.5%	Score 33.8	DB 3	Length 20235
Best Local	Similarity	53.0%	Pred. No. 1.9		
Matches	96	Conservative	0	Mismatches	82
				Indels	3
				Gaps	1
QY	21	CAGCAGCCGTGGCCCGAACAACAGCTCAGAGACGTGCGCGAGAGATGGCCGACTTGGCCG	80		
Db	3654	CGGAGAGATCAAGCAAGTGGCAGCCGAGAGATTTGCCGAGAGATGGCGGCACAGACCGC	3713		
QY	81	CATGAAACCGTGCTGCTGTAACCTGTGTGGCCCTTGCATGCACGAAAGGGGAATGTTTC	140		
Db	3714	GTGTGTCGCCCTGTGCGAGCAAGTCTCT--CGGCGACCGGGGCCGTGCGAGCGGAGTCTTA	3770		
QY	141	CTGCCCGTGTAATGCGGCTCACTACAGAGGATTCCTACAGGACCACTGCGAGAGCCGACAG	200		
Db	3771	CTGCTCGTGGTGGCCGGGATCTGGGGCGGCGCCGAGTGGCCGCTTACCGGCGCAGGGCAG	3830		
QY	201	C 201			
Db	3831	C 3831			

```

1 APPLICATION NUMBER: US/08/676,974
2 FILING DATE:
3 CLASSIFICATION: 530
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Osman Ph.D., Richard A
6 REGISTRATION NUMBER: 36,627
7 REFERENCE/DOCKET NUMBER: UCB96-055
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: (415)343-4341
10 TELEFAX: (415)343-4342
11 INFORMATION FOR SEQ ID NO: 5:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 2277 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: double
16 TOPOLOGY: linear
17 MOLECULE TYPE: CDNA
18 US-08-676-974-5
19
20 Query Match 10.3%; Score 33.2; DB 1; Length 2277;
21 Best Local Similarity 67.1%; Pred.No.1.8;
22 Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0
23
24 Oy 176 GCACCACTGAGAGAGGCGACGACCTGGCCGACACACAGCTCAAGACGTGGCGAGA 235
25 || ||| ||||| ||| ||| || | | ||| |||
26 Db 842 GCGACCACTGAGAGAGGCGACGACCTGGAGAGAGACGACACGATCGACGAGCGGAGG 901
27 || ||| ||| ||| ||| ||| ||| ||| |||
28 Oy 236 AGATGGCCGA 245
29 || ||||| |
30 Db 902 AGCTGGCCCA 911
31
32 RESULT 12
33 US-09-098-487-5
34 ; Sequence 5, Application US/09098487
35 ; Patent No. 5917025
36
37 GENERAL INFORMATION:
38 APPLICANT: COLLINS, Kathleen
39 TITLE OF INVENTION: Human Telomerase
40 NUMBER OF SEQUENCES: 11
41 CORRESPONDENCE ADDRESSES:
42 ADDRESSEE: Science & Technology Law Group
43 STREET: 268 Bush Street, Suite 3200
44 CITY: San Francisco
45 STATE: CA
46 COUNTRY: USA
47 ZIP: 94104
48
49 COMPUTER READABLE FORM:
50 MEDIUM TYPE: floppy disk
51 COMPUTER: IBM PC compatible
52 OPERATING SYSTEM: PC-DOS/MS-DOS
53 SOFTWARE: PatentIn Release #1.0, Version #1.30
54 CURRENT APPLICATION DATA:
55 APPLICATION NUMBER: US/09/098,487
56 FILING DATE:
57
58 CLASSIFICATION:
59 ATTORNEY/AGENT INFORMATION:
60 NAME: Osman Ph.D., Richard A
61 REGISTRATION NUMBER: 36,627
62 REFERENCE/DOCKET NUMBER: UCB96-055
63 TELECOMMUNICATION INFORMATION:
64 TELEPHONE: (415)343-4341
65 TELEFAX: (415)343-4342
66 INFORMATION FOR SEQ ID NO: 5:
67 SEQUENCE CHARACTERISTICS:
68 LENGTH: 2277 base pairs
69 TYPE: nucleic acid
70 STRANDEDNESS: double
71 TOPOLOGY: linear
72 MOLECULE TYPE: CDNA
73 US-09-098-487-5
74
75 Query Match 10.3%; Score 33.2; DB 2; Length 2277;

```



```

      STREET: 3000 K Street, N.W.  

      CITY: Washington, D.C.  

      COUNTRY: USA  

      ZIP: 20007-5109  

      COMPUTER READABLE FORM:  

      MEDIUM TYPE: Floppy disk  

      COMPUTER: IBM PC compatible  

      OPERATING SYSTEM: PC-DOS/MS-DOS  

      SOFTWARE: Patent In Release #1.0, Version #1.25  

      CURRENT APPLICATION DATA:  

      APPLICATION NUMBER: US/08/173,508  

      FILING DATE: 23-DEC-1993  

      CLASSIFICATION: 435  

      ATTORNEY/AGENT INFORMATION:  

      NAME: BENT, Stephen A.  

      REGISTRATION NUMBER: 29,768  

      REFERENCE/DOCKET NUMBER: 18740/125/CACO  

      TELECOMMUNICATION INFORMATION:  

      TELEPHONE: 202 672 5300  

      TELEFAX: 202 672 5399  

      TELEX: 904136  

      INFORMATION FOR SEQ ID NO: 5:  

      SEQUENCE CHARACTERISTICS:  

      LENGTH: 1777 base pairs  

      TYPE: nucleic acid  

      STRANDEDNESS: double  

      TOPOLOGY: linear  

      FEATURE:  

      NAME/KEY: CDS  

      LOCATION: 190..1731  

      FEATURE:  

      NAME/KEY: misc.feature  

      LOCATION: 190..192  

      OTHER INFORMATION: /note= "Met at position 1  

        OTHER INFORMATION: represents fMet"  

US-08-173-508-5

```

---

```

Query Match          10.2%; Score 32.6; DB 1; Length 1777;  

Best Local Similarity 63.3%; Pred. No. 2.5;  

Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY    165 GGGATCCTCAAGCACCCACTGCGAGGAGGCCAGCAGCCTTGCGCGAACAAGTCAAGA 224
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1539 GGGGAGACCCGCGCCACGCCTACGAGGGCGCCCCAGAGATGGCGACGAGCTGGGCAAAGA 1598
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY    225 CGTCGGCGAGAGAAGATGGCC 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1599 CGTCGGCGCTGTGCTGACC 1617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Search completed: June  2, 2003, 06:54:04
Job time : 57 secs
```







RESULT 2  
US-09-815-242-7812  
; Sequence 7812, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7812  
; LENGTH: 1431  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1431)  
US-09-815-242-7812

Query Match 11.2%; Score 35.8; DB 10; Length 1431;  
Best Local Similarity 45.5%; Pred. No. 0.036;  
Matches 127; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 15 GAGGCCAGCAGCTGGCCGGAACACACAGCTCAAGAGAGTGGCGGAGAGTGGCGGACTT 74  
DB 924 GGGGGTGTAGCTGCGAGCGAAGCCGACCGCGACCGCCGACATCCGACCCCTGCGCAAT 983  
QY 75 GGGCGCATGGAACCGGTGCTGTGAAGTCTGTGCGCTGCGCTGCATGCAGAGAGGGA 134  
DB 984 GACCAAGAGCGCATCGAGCTGTGAAGACATCCGAGCGGCTTCTTCTGCGAGGTGA 1043  
QY 135 TGTTCCTGCGCCGTGTGATCGCGTCACTACAGGATCTTCAGGCAAGCCACTGCGAGAGG 194  
DB 1044 GGGCGGTGATGATGACAGAGGAGGACATCCGCGATCCGTGCGGCAATGCGGCGAGAC 1103  
QY 195 CAGCAGCTGCGCGGACAGACAACTCAAGAGAGTGGCGGAGAGAGTGGCGGCGG 254  
DB 1104 CGTGCACCTGAGCAGACCGCGTGTGCGAGAGGCGCTTGTGCGAAGGCGGATGCGAGAC 1163  
QY 255 CATGGAACCGTGTGTGTGACTGCTGTGCGGCTGCGCA 293  
DB 1164 CCGTGTGATGTGACCGCGGACGACCGCCACTCCAGCA 1202

RESULT 3  
US-10-166-087-17  
; Sequence 17, Application US/10166087  
; Publication No. US2003007767A1  
; GENERAL INFORMATION:

; APPLICANT: Ecopia Biosciences Inc.  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin  
; FILE REFERENCE: 3014-205  
; CURRENT APPLICATION NUMBER: US/10/166,087  
; CURRENT FILING DATE: 2002-06-11  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 17  
; LENGTH: 2295  
; TYPE: DNA  
; ORGANISM: Streptomyces refuineus subspecies thermotolerans  
US-10-166-087-17

Query Match 10.8%; Score 34.6; DB 9; Length 2295;  
Best Local Similarity 58.1%; Pred. No. 0.093;  
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 177 CACCCACTGCGAGAGGCGCAGCAGCTGCGCGAACAACAGCTCAAGAGAGTGGCGGAGAA 236  
DB 843 CGCTCCAGAGATCGACGCGAAGAACATCGCGACTACGCGCGCATGAGGTGAGCGAACT 902  
QY 237 GATGCGCGACTTGGCGCGCATGGAACCGTGTGTGAACTGCT 281  
DB 903 GATCACCAGCTGGGCGCATGACGACCGGTGGCGAACCCT 947

RESULT 4  
US-10-166-087-1  
; Sequence 1, Application US/10166087  
; Publication No. US2003007767A1  
; GENERAL INFORMATION:  
; APPLICANT: Ecopia Biosciences Inc.  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin  
; FILE REFERENCE: 3014-205  
; CURRENT APPLICATION NUMBER: US/10/166,087  
; CURRENT FILING DATE: 2002-06-11  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 32539  
; TYPE: DNA  
; ORGANISM: Streptomyces refuineus subspecies thermotolerans  
US-10-166-087-1

Query Match 10.8%; Score 34.6; DB 9; Length 32539;  
Best Local Similarity 58.1%; Pred. No. 0.13;  
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 177 CACCCACTGCGAGAGGCGCAGCAGCTGCGCGAACAACAGCTCAAGAGAGTGGCGGAGAA 236  
DB 11359 CGCTCCAGAGATCGACGCGAAGAACATCGCGACTACGCGCGCATGAGGTGAGCGAACT 11418  
QY 237 GATGCGCGACTTGGCGCGCATGGAACCGTGTGTGAACTGCT 281  
DB 11419 GATCACCAGCTGGGCGCATGACGACCGGTGGCGAACCCT 11463

RESULT 5  
US-10-044-090-167  
; Sequence 167, Application US/10044090  
; Patent No. US20020137081A1  
; GENERAL INFORMATION:  
; APPLICANT: Olga Bandman  
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
; FILE REFERENCE: PA-0028 US  
; CURRENT APPLICATION NUMBER: US/10/044,090  
; CURRENT FILING DATE: 2002-01-09

```

: NUMBER OF SEQ ID NOS: 850
: SOFTWARE: PERL Program
: SEQ ID NO 167
: LENGTH: 3759
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: OTHER INFORMATION: Incyte ID No. US20020137081A1 406362.7
: LOCATION: 73
: OTHER INFORMATION: a, t, c, g, or other
: US-10-044-090-167

```

Query Match	10.4%	Score 33.4	DB 12	Length 3759
Best Local Similarity	53.0%	Pred. No	0.24	
Matches	98	Conservative	0	Mismatches 81
				Indels 6
				Gaps 1

QY	134	ATGTTTCTGCGCCCTTATCGCGTCACATCAACAGGAGTCTTCAGGGACCCACTGCGAGGAG	193
Db	2774	ATGTTGCTGCGCCAGTGAATGTTCCAAAGACTGCGCCAAAGATCAACGCGGCCCGAAGAG	2833
QY	194	CCAGCAGCCTGTGCC-----GAAACAAGACTCAAGAGAGCTGCGGAGAGATGGCCGACT	247
Db	2834	CCCTCATCTGTGGGCTTCATGGCCGCGCTCCGAGAGAACCCGTCGGCAGGAGCAGTGTGACG	2899
QY	248	TGCGCGCGCATGGAAAACGCTGCTGTCTGTACTTCGTGTGCGCCTGCCATGCACGAAAGGGA	307
Db	2894	TGATTCAGATCAAAAGCTGAGCTGTATACATCAACGAGGAGCACTTGGCCAAAGGGGAGAGGCA	2953
QY	308	ATGTT 312	
Db	2954	GGGTT 2958	

```

RESULT 6
US-10-175-523-48
: Sequence 48, Application US/10175523
: Publication No. US20030096264A1
: GENERAL INFORMATION:
: APPLICANT: Brockman, Jeffrey
: APPLICANT: Evans, David
: APPLICANT: Hook, Derek
: APPLICANT: Klimczak, Leszek
: APPLICANT: Laeng, Pascal
: APPLICANT: Palfreyman, Michael
: APPLICANT: Rajan, Prithi
: TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)

```

```

? CURRENT APPLICATION NUMBER: US/10/175,523
? CURRENT FILING DATE: 2002-06-18
? PRIOR APPLICATION NUMBER: US 60/299,151
? PRIOR FILING DATE: 2001-06-18
? PRIOR APPLICATION NUMBER: US 60/317,828
? PRIOR FILING DATE: 2001-09-07
? PRIOR APPLICATION NUMBER: US 60/325,150
? PRIOR FILING DATE: 2001-09-25
? PRIOR APPLICATION NUMBER: US 60/333,047
? PRIOR FILING DATE: 2001-11-14
? PRIOR APPLICATION NUMBER: US 60/349,936
? PRIOR FILING DATE: 2002-01-18
? PRIOR APPLICATION NUMBER: US 60/361,834
? PRIOR FILING DATE: 2002-03-04
? NUMBER OF SEQ ID NOS: 197
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 48
? LENGTH: 2838
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-10-175-523-48

```

Query Match	10.38;	Score 33;	DB 9;	Length 2838;
Best Local Similarity	45.38;	Pred. No. 0.31;		

	Matches	120;	Conservative	0;	Mismatches	145;	Indels	0;	Gaps	0;
QY	27	CCTGGCCGACACAACTGCTCAAGGACGTGCGCGAAGAAAGTGGCCACTTGGCGCGCATGA	86							
Db	864	CGTCTGGGGGTCAAGATCAAGAGAGGGCCCGGTGAAGCGCGCGGGGCCCGCGGGGGG	923							
QY	87	AACCGTGTCTCTAACACTGCTGTGCGCGCTGCGCATGACAGAAAGGGGAATGTTTCTGTGCC	146							
Db	924	CGCGGCGGCGCGGTGGGCGAGTTCATGTTCGACGCTGTGCAAGGAGGAGTACGCGGACCC	983							
QY	147	GTTGATGCGGCTACTACAGGGATCTCTAGGAGACCACATCGCAGAGAGGCGACAGCGTGGC	206							
Db	984	GTTGCGCTGGCGAGCACAAATGTCGCGCATGTGCGTGTGGAGTACCGGTGTCCCGA	1043							
QY	207	CGAACACAAAGCTCAAGGACGTGCGCGAGAGATGGCCGACTTGGCGCATGGAACCGT	266							
Db	1044	GTTGGCCCAAGGTTTTCACGTGCGCGGCGACAACTGGCTTCGACCGCGCTGGACAAAC	1103							
QY	267	GCTGCTGAACTCGATGGTGGCGCTTG	291							
Db	1104	GGGCGCGCGCGCGCGCGCGCGCGC	1128							

```

RESULT 7
US-09-954-456-700
; Sequence 700, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456

```

```

? PRIOR APPLICATION NUMBER: US/60/233,617
? PRIOR FILING DATE: 2000-09-18
? PRIOR APPLICATION NUMBER: US/60/234,052
? PRIOR FILING DATE: 2000-09-20
? PRIOR APPLICATION NUMBER: US/60/234,923
? PRIOR FILING DATE: 2000-09-25
? PRIOR APPLICATION NUMBER: US/60/235,134
? PRIOR FILING DATE: 2000-09-25
? PRIOR APPLICATION NUMBER: US/60/235,637
? PRIOR FILING DATE: 2000-09-26
? PRIOR APPLICATION NUMBER: US/60/235,638
? PRIOR FILING DATE: 2000-09-26
? PRIOR APPLICATION NUMBER: US/60/235,711
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: US/60/235,720
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: US/60/235,840
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: US/60/235,863
? PRIOR FILING DATE: 2000-09-27
? NUMBER OF SEQ ID NOS: 2276
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 700
? LENGTH: 2838
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-954-456-700

```

Query Match	10.3%;	Score 33;	DB 10;	Length 2838;
Best Local Similarity	45.3%;	Pred. No. 0.31;		

27 CCTGGCCGACACAGCTCAAGCAGTGGCGGAGCAAGATGGCCGACTTGGCGGCATGGA 86

Db 864 CGTGGTGGGGCTCAAGATCAAGAGAGGGCCCCGGTGGAGAGCGCGGGGGCCGGCGGGGGGG 923

Db 924 CGCGCGCGCGCGCTGGGCGAGTTCATCTGCCAGCTGTGCAGAGGAGGAGTACGCCGACCC 983





```
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086US1/10026309
; CURRENT FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 72
; LENGTH: 5502
; TYPE: DNA
; ORGANISM: Human
US-10-021-955-72
```

```
Query Match 10.3%; Score 33; DB 9; Length 5502;
Best Local Similarity 57.1%; Pred. No. 0.34;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 18 GCCACGACGCTGGCCGGAACAGCTCAAGACGTCGGCGAGAGATGGCCGACTTGGC 77
DB 593 GCCCGGACCGGTGTGCTGCTACGAGATCAAGGCCCGCGGCCAAGGTGGCCAGCTGCT 652
QY 78 GCCGATGGAACCGTGTCTGTGAAGTCTGTCGTCGGCTGGCCATGC 122
DB 653 ACGGCTGTAGCCCGCGCCCGCCCTGAGACTGCCCAAGCATCC 697
```

```
RESULT 14
US-10-021-955-75
; Sequence 75, Application US/10021955
; Publication No. US20030039987A1
; GENERAL INFORMATION:
; APPLICANT: Lupski, James R
; APPLICANT: Boerkoel, Cornelius F
; APPLICANT: Takashima, Hiroshi
; TITLE OF INVENTION: Defects in Periaxin Associated with Myelinopathies
; FILE REFERENCE: P02086US1/10026309
; CURRENT FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: US 60/255,217
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 75
; LENGTH: 5502
; TYPE: DNA
; ORGANISM: Human
US-10-021-955-75
```

```
Query Match 10.3%; Score 33; DB 9; Length 5502;
Best Local Similarity 57.1%; Pred. No. 0.34;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 18 GCCACGACGCTGGCCGGAACAGCTCAAGACGTCGGCGAGAGATGGCCGACTTGGC 77
DB 593 GCCCGGACCGGTGTGCTGCTACGAGATCAAGGCCCGCGGCCAAGGTGGCCAGCTGCT 652
QY 78 GCCGATGGAACCGTGTCTGTGAAGTCTGTCGTCGGCTGGCCATGC 122
DB 653 ACGGCTGTAGCCCGCGCCCGCCCTGAGACTGCCCAAGCATCC 697
```

```
RESULT 15
US-09-975-719-160
; Sequence 160, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
```

```
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 160
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-160
```

```
Query Match 10.2%; Score 32.8; DB 9; Length 1299;
Best Local Similarity 54.0%; Pred. No. 0.33;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
```

```
QY 144 CCCGTGATCGCCTCACTACAGGATCTCTAGGACCCCACTGCGAGAGGCCAGGACCT 203
DB 885 CCTCTCCCGGGGTATCCGAGGGTTACTACGAGAACCACTCTCTGATCTGAGCT 944
QY 204 GGCCGAACACAGCTCAAGGACGTCGGCGAGAGATGGCCGACTTGGCCGATGGAAC 263
DB 945 GAGCGACGGCAAGATCAAGCGCAACCGGATTCATGACGTCCTCCAGCAATTCGGCGC 1004
QY 264 CGTG 267
DB 1005 CCTG 1008
```

```
Search completed: June 2, 2003, 07:36:45
Job time : 138 secs
```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:23:57 ; Search time 35 Seconds

(without alignments)  
445.438 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 613  
Sequence: 1 MTHCEASSLAEHLKDYRE.....HARKGNVSCPSAWSHQFEK 117

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	613	100.0	117	23	AAU97552
2	607	99.0	117	23	AAU97560
3	605	98.7	117	23	AAU97553
4	605	98.7	117	23	AAU97556
5	605	98.7	117	23	AAU97557
6	603	98.7	117	23	AAU97558
7	603	98.4	117	23	AAU97559
8	601	98.0	117	23	AAU97555
9	594.5	97.0	118	23	AAU97554
10	290	47.3	144	23	AAU97551

11	276	45.0	144	20	AAV01816
12	257	41.9	159	15	AAR49668
13	125	20.4	135	15	AAR49669
14	123	20.1	135	15	AAR49670
15	92	15.0	138	23	ABP38137
16	88.5	14.4	296	22	ABM56069
17	88.5	14.4	296	23	ABB72269
18	88.5	14.4	299	21	AAV76014
19	88.5	14.4	299	21	AAV76017
20	88.5	14.4	299	22	ABM55953
21	88.5	14.4	299	22	ABM56016
22	88.5	14.4	299	23	ABB72153
23	88.5	14.4	299	23	ABB72156
24	88	14.4	130	23	ABP26323
25	88	14.4	130	23	ABP28475
26	81	13.2	814	21	AAV29830
27	81	13.2	823	21	AAV29829
28	81	13.2	880	21	AAV29828
29	78.5	12.8	268	22	ABG15429
30	78.5	12.8	347	22	ABG12165
31	78.5	12.8	264	22	ABG02343
32	77	12.6	340	22	ABG06450
33	76.5	12.5	475	22	AAV81188
34	76.5	12.5	671	23	ABM49703
35	76	12.4	1037	22	ABM49499
36	75.5	12.3	799	22	ABM58486
37	75	12.2	2442	21	AAV77575
38	73.5	12.0	510	22	ABG11601
39	73	11.9	352	21	AAV57596
40	72.5	11.8	225	22	ABM61600
41	72.5	11.8	933	22	ABM60762
42	72	11.7	264	22	AAV80130
43	72	11.7	400	22	AAV9146
44	72	11.7	999	21	AAV73362
45	72	11.7	1388	19	AAV56475

#### ALIGNMENTS

RESULT 1	
AAU97552	
ID	AAU97552 standard; Protein: 117 AA.
XX	
AC	AAU97552;
XX	
DT	13-AUG-2002 (first entry)
XX	
DE	Synthetic Merr chelon variant.
XX	
KW	Mercuric ion; contaminated soil; ground water; hydroponic solution;
KW	irrigation water; waste stream; contaminated aqueous medium;
KW	biological fluid; gastrointestinal tract; chelon protein;
KW	enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;
KW	heavy metal binding protein; Merr.
XX	
OS	Synthetic.
XX	
PN	W0200230962-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US31819.
XX	
PR	12-OCT-2000; 2000US-240465P.
XX	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.
XX	
PI	Summers AO, Cagulat JJ;
XX	
DR	WPI, 2002-435437/46.
XX	
DR	N-PSDB; ABR32207, ABR52213.
XX	

Merr protein which  
protein product of  
protein product of  
protein product of  
Staphylococcus epi  
Skin cell protein,  
Merr protein iso  
Merr protein iso  
Merr EGF family  
Merr EGF family  
Skin cell protein,  
Skin cell protein,  
Merr protein iso  
Merr protein iso  
Streptococcus poly  
Streptococcus poly  
Arabidopsis thalia  
Arabidopsis thalia  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Listeria monocytoc  
Novel human diagno  
Drosophila melanog  
Human cytoskeletal  
Novel human diagno  
Merr Wnt-3a prot  
Novel human diagno  
Drosophila melanog  
Human protein SBO  
Human protein SBO  
Human clone 2674047  
protein with Rho p

PT Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
Claim 2; Page 20; 42pp; English.

The present invention relates to a new non-naturally occurring recombinant DNA molecule comprising a sequence encoding a chelon protein which binds mercuric ions. The invention is useful for recombinantly producing a protein in a host-cell, by infecting or transforming a host cell capable of expressing a chelon coding sequence with a vector comprising a promoter active in the host cell operably linked to a coding region for the protein to produce a recombinant host cell and culturing the recombinant host cell under conditions, where DNA is expressed. The nucleic acid encoding the chelon protein is useful for binding divalent mercuric ions, to take up, sequester and concentrate the heavy metal ions from contaminated soil, ground water, hydroponic solutions or irrigation water of waste streams. The DNA of the invention, when immobilised onto a solid support, is useful for concentrating heavy metal ions from contaminated environment waste streams or contaminated aqueous medium including biological fluids. The nucleic acid, when recombinantly expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic), is suitable for use in the in vivo sequestration and elimination of mercuric ion from gastrointestinal tracts of animals or humans exposed to toxic metal ions such as mercury and/or cadmium. The molecules of the invention are also useful in water treatment resins. The nucleic acid of the invention is highly specific and binds divalent cation such as mercury or cadmium with high affinity. The present amino acid sequence represents the synthetic MERR chelon variant protein of the invention. This sequence is one of the heavy metal binding proteins termed Chelons of the invention.

Note: ABRK32207 encodes only amino acids 1-107 of this protein.

SO	Sequence	117	AA;
	Query Match	100.0%	Score 613; DB 23;
	Best Local Similarity	100.0%	Pred. NO. 3.2e-53;
	Matches 117; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

QY	1	MT	HC	EA	SL	AE	HT	KL	KD	VR	KE	MA	DI	AR	ET	VL	SE	VC	AC	HA	R	GN	VS	CP	LS	SI	SS	ST	60
QY	1	MT	HC	EA	SL	AE	HT	KL	KD	VR	KE	MA	DI	AR	ET	VL	SE <td>VC</td> <td>AC</td> <td>HA</td> <td>R</td> <td>GN</td> <td>VS</td> <td>CP</td> <td>LS</td> <td>SI</td> <td>SS</td> <td>ST</td> <td>60</td>	VC	AC	HA	R	GN	VS	CP	LS	SI	SS	ST	60
Db	1	MT	HC	EA	SL	AE	HT	KL	KD	VR	KE	MA	DI	AR	ET	VL	SE <td>VC</td> <td>AC</td> <td>HA</td> <td>R</td> <td>GN</td> <td>VS</td> <td>CP</td> <td>LS</td> <td>SI</td> <td>SS</td> <td>ST</td> <td>60</td>	VC	AC	HA	R	GN	VS	CP	LS	SI	SS	ST	60
QY	61	HC	EA	SL	AE	HT	KL	KD	VR	KE	MA	DI	AR	ET	VL	SE <td>VC</td> <td>AC</td> <td>HA</td> <td>R</td> <td>GN</td> <td>VS</td> <td>CP</td> <td>LS</td> <td>SI</td> <td>SS</td> <td>ST</td> <td>60</td>	VC	AC	HA	R	GN	VS	CP	LS	SI	SS	ST	60	
Db	61	HC	EA	SL	AE	HT	KL	KD	VR	KE	MA	DI	AR	ET	VL	SE <td>VC</td> <td>AC</td> <td>HA</td> <td>R</td> <td>GN</td> <td>VS</td> <td>CP</td> <td>LS</td> <td>SI</td> <td>SS</td> <td>ST</td> <td>60</td>	VC	AC	HA	R	GN	VS	CP	LS	SI	SS	ST	60	

## RESULT 2

ID AU97560 standard; Protein; 117 AA.

AAU97560, AC

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #8.

KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.

OS Synthetic.

PN W0200230962-A2.

PD 18-APR-2002

12-OCT-2001; 2001WO-US31819.

PR 12-OCT-2000; 2000US-240465P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC  
PA  
XX  
XX Summers AO, Caguiat JJ;  
PI  
XX  
XX  
DR WPI; 2002-435437/46.

PT Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
XX  
PS Claim 4; Page 22; 42pp; English.

The present invention relates to a new non-naturally occurring recombinant DNA molecule comprising a sequence encoding a chelon protein which binds mercuric ions. The invention is useful for recombinantly producing a protein in a host cell, by infecting or transforming a host cell capable of expressing a chelon coding sequence with a vector comprising a promoter active in the host cell operably linked to a coding region for the protein to produce a recombinant host cell and culturing the recombinant host cell under conditions, where DNA is expressed. The nucleic acid encoding the chelon protein is useful for binding divalent mercuric ions, to take up, sequester and concentrate the heavy metal ions from contaminated soil, ground water, hydroponic solutions or irrigation water of waste streams. The DNA of the invention, when immobilised onto a solid support, is useful for concentrating heavy metal ions from contaminated environment waste streams or contaminated aqueous medium including biological fluids. The nucleic acid, when recombinantly expressed in enteric bacteria (which are nonoxigenic and nonpathogenic), is suitable for use in the *in vivo* sequestration and elimination of mercuric ion from gastrointestinal tracts of animals or humans exposed to toxic metal ions such as mercury and/or cadmium. The molecules of the invention are also useful in water treatment resins. The nucleic acid of the invention is highly specific and binds divalent cation such as mercury or cadmium with high affinity. The present amino acid sequence represents one of a collection (AA97553-AA97560) of synthetic cadmium/mercury ion binding chelon proteins of the invention. This sequence is one of the heavy metal binding proteins termed chelons of the invention.

Sequence	117 AA;				
Query Match	99.0%;	Score 607;	DB 23;	Length 117;	
Best Local Similarity	99.1%;	Pred. No. 1.3e-52;			
Matches 116; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;	

QY	1	MHCEBASSLAHKKLDKDYERKADLARMTYVLSLYCACHARKGNVSCPLILNLOSSST	60
QY	1	MHCEBASSLAHKKLDKDYERKADLARMTYVLSLYCACHARKGNVSCPLILNLOSSST	60
Db	1	MHCEBASSLAHKKLDKDYERKADLARMTYVLSLYCACHARKGNVSCPLILNLOSSST	60
QY	61	HCEBASSLAHKKLDKDYERKADLARMTYVLSLYCACHARKGNVSCPSMWSHPFK	117
Db	61	HCEBASSLAHKKLDKDYERKADLARMTYVLSLYCACHARKGNVSCPSMWSHPFK	117

RESULT 3  
AAU97553

AAU97553; .AC

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #1.

KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
KW Irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.

OS Synthetic.



PN WO200230962-A2.  
XX 18-APR-2002.  
XX 12-OCT-2001; 2001WO-US31819.  
XX 12-OCT-2000; 2000US-240465P.  
PR 12-OCT-2000; 2000US-240465P.  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX Summers AO, Caguiat JI;  
PI WPI; 2002-435437/46.  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from a  
PT contaminated soil, water, aqueous medium including biological fluids -  
PS Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AA097553-AA097560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.

CC Sequence 117 AA:

Query Match 98.7%; Score 605; DB 23; Length 117;  
Best Local Similarity 98.3%; Pred. No. 2e-52;  
Matches 115; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTHCEASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPLASLOGSSGT 60  
DB 1 MTHCEASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPLASLOGSSGT 60

QY 61 HCEBASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPSAMSHPOFEK 117  
DB 61 HCEBASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPSAMSHPOFEK 117

RESULT 4  
AA097556 standard; Protein: 117 AA.

XX AA097556;

XX 13-AUG-2002 (first entry)

XX Synthetic cadmium/mercury ion binding chelon protein #4.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;

KW irrigation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein.

OS Synthetic.

PN WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31819.

PR 12-OCT-2000; 2000US-240465P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Summers AO, Caguiat JI;

PI WPI; 2002-435437/46.

XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PT contaminated soil, water, aqueous medium including biological fluids -  
PS Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AA097553-AA097560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.

CC Sequence 117 AA:

Query Match 98.7%; Score 605; DB 23; Length 117;  
Best Local Similarity 98.3%; Pred. No. 2e-52;  
Matches 115; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTHCEASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPLASLOGSSGT 60  
DB 1 MTHCEASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPLASLOGSSGT 60

QY 61 HCEBASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPSAMSHPOFEK 117  
DB 61 HCEBASSLAHEHKLKDYREKMDLARMETVLSLYCACHARKGNVSCPSAMSHPOFEK 117

RESULT 5  
AA097557 standard; Protein: 117 AA.

XX AA097557;

AAU97557; 13-AUG-2002 (first entry)  
Synthetic cadmium/mercury ion binding chelon protein #5.  
Mercuric ion; contaminated soil; ground water; hydroponic solution;  
irrigation water; waste stream; contaminated aqueous medium;  
biological fluid; gastrointestinal tract; chelon protein;  
enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
heavy metal binding protein.  
Synthetic.  
WC200230962-A2.  
18-APR-2002.  
12-OCT-2001; 2001WO-US31819.  
12-OCT-2000; 2000US-240465P.  
(UYGE-) UNIV GEORGIA RES FOUND INC.  
Summers AO, Caguiat JI;  
WPI: 2002-435437/46.  
Novel non-naturally occurring recombinant DNA molecule encoding a  
chelon protein useful for binding divalent cation mercury from  
contaminated soil, water, aqueous medium including biological fluids -  
Claim 4; Page 22; 42pp; English.

The present invention relates to a new non-naturally occurring  
recombinant DNA molecule comprising a sequence encoding a chelon protein  
which binds mercuric ions. The invention is useful for recombinantly  
producing a protein in a host-cell, by infecting or transforming a host  
cell capable of expressing a chelon coding sequence with a vector  
comprising a promoter active in the host cell operably linked to a coding  
region for the protein to produce a recombinant host cell and culturing  
the recombinant host cell under conditions, where DNA is expressed.  
The nucleic acid encoding the chelon protein is useful for binding  
divalent mercuric ions, to take up, sequester and concentrate the heavy  
metal ions from contaminated soil, ground water, hydroponic solutions or  
irrigation water of waste streams. The DNA of the invention, when  
immobilised onto a solid support, is useful for concentrating heavy metal  
ions from contaminated environment waste streams or contaminated  
aqueous medium including biological fluids. The nucleic acid, when  
recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
nonpathogenic), is suitable for use in the in vivo sequestration and  
elimination of mercuric ion from gastrointestinal tracts of animals or  
humans exposed to toxic metal ions such as mercury and/or cadmium. The  
molecules of the invention are also useful in water treatment resins.  
The nucleic acid of the invention is highly specific and binds divalent  
cation such as mercury or cadmium with high affinity. The present amino  
acid sequence represents one of a collection (AAU97553-AAU97560) of  
synthetic cadmium/mercury ion binding chelon proteins of the invention.  
This sequence is one of the heavy metal binding proteins termed chelons  
of the invention.

Sequence 117 AA;

Query Match 98.7%; Score 605; DB 23; Length 117;  
Best Local Similarity 98.3%; Pred. No. 2e-53; Indels 0; Gaps 0;  
Matches 115; Conservative 2; Mismatches 0;

1 MTHCEASSLAHKKLDYREKMDLARMEVTVLSLVACAHARKGNVSCPLIASLOGSSGT 60  
|||||  
1 MTHCEASSLAHKKLDYREKMDLARVEVTVLSLVACAHARKGNVSCPLIASLOGSSST 60  
|||||

61 HCEASSLAHKKLDYREKMDLARMEVTVLSLVACAHARKGNVSCPSAWSHPOFEK 117  
|||||

Db 61 HCEBMSLAJEHKLKQVREKMDLAVETVLSLVCACAHARGVNSCPASMSHPQEK 117

RESULT 6  
AAU97558  
ID AAU97558 standard; Protein: 117 AA.  
XX AC AAU97558;  
XX DT 13-AUG-2002 (first entry)  
XX DE Synthetic cadmium/mercury ion binding chelon protein #6.  
XX KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
XX KW Irrigation water; waste stream; contaminated aqueous medium;  
XX KW biological fluid; gastrointestinal tract; chelon protein;  
XX KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
XX KW heavy metal binding protein.  
XX OS Synthetic.  
XX PN WO200230962-A2.  
XX PR 18-APR-2002.  
XX PF 12-OCT-2001; 2001WO-US31819.  
XX PR 12-OCT-2000; 2000US-240465P.  
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX PI Summers AO, Caquiat JT;  
XX DR WPI: 2002-435437/46.  
XX PT Novel non-naturally occurring recombinant DNA molecule encoding a  
XX PT chelon protein useful for binding divalent cation mercury from  
XX PT contaminated soil, water, aqueous medium including biological fluids -  
XX PS Claim 4; Page 22; 42pp; English.

The present invention relates to a new non-naturally occurring recombinant DNA molecule comprising a sequence encoding a chelon protein which binds mercuric ions. The invention is useful for recombinantly producing a protein in a host-cell, by infecting or transforming a host cell capable of expressing a chelon coding sequence with a vector comprising a promoter active in the host cell operably linked to a coding region for the protein to produce a recombinant host cell and culturing the recombinant host cell under conditions, where DNA is expressed. The nucleic acid encoding the chelon protein is useful for binding divalent mercuric ions, to take up, sequester and concentrate the heavy metal ions from contaminated soil, ground water, hydroponic solutions or irrigation water of waste streams. The DNA of the invention, when immobilised onto a solid support, is useful for concentrating heavy metal ions from contaminated environment waste streams or contaminated aqueous medium including biological fluids. The nucleic acid, when recombinantly expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic), is suitable for use in the in vivo sequestration and elimination of mercuric ion from gastrointestinal tracts of animals or humans exposed to toxic metal ions such as mercury and/or cadmium. The molecules of the invention are also useful in water treatment resins. The nucleic acid of the invention is highly specific and binds divalent cation such as mercury or cadmium with high affinity. The present amino acid sequence represents one of a collection (AAU97553-AAU97560) of synthetic cadmium/mercury ion binding chelon proteins of the invention. This sequence is one of the heavy metal binding proteins termed chelons of the invention.

Sequence 117 AA;  
SQ

Query Match 98.7%; Score 605; DB 23; Length 117;  
Best Local Similarity 98.3%; Pred. No. 26-52;  
Matches 115; Conservative 2; Mismatches 0; Indels 0; Gaps 0

QY 1 MTHCEASSLAHKLKDVREKMDLARMETVLSLVACAHARGNVSCPLIASIQSSGT 60  
 |||||  
 DB 1 MTHCEASSLAHKLKDVREKMDLARIETVLSLVACAHARGNVSCPLIASIQSSGT 60  
 |||||  
 QY 61 HCEASSLAHKLKDVREKMDLARMETVLSLVACAHARGNVSCPSAMSHPOFEK 117  
 |||||  
 DB 61 HCEASSLAHKLKDVREKMDLARIETVLSLVACAHARGNVSCPSAMSHPOFEK 117  
 |||||  
 RESULT 7  
 AAU97555  
 ID AAU97555 standard; Protein: 117 AA.  
 AC AAU97555;  
 DT 13-AUG-2002 (first entry)  
 DE Synthetic cadmium/mercury ion binding chelon protein #7.  
 KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
 KW irrigation water; waste stream; contaminated aqueous medium;  
 KW biological fluid; gastrointestinal tract; chelon protein;  
 KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
 KW heavy metal binding protein.  
 OS Synthetic.  
 PN MO200230962-A2.  
 PD 18-APR-2002.  
 PF 12-OCT-2001; 2001WO-US31819.  
 PR 12-OCT-2000; 2000US-240465P.  
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 PI Summers AO, Caguiat JJ;  
 PS MPI: 2002-435437/46.  
 PT Novel non-naturally occurring recombinant DNA molecule encoding a  
 chelon protein useful for binding divalent cation mercury from  
 contaminated soil, water, aqueous medium including biological fluids -  
 Claim 4; Page 22; 42pp; English.  
 CC The present invention relates to a new non-naturally occurring  
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
 CC which binds mercuric ions. The invention is useful for recombinantly  
 CC producing a protein in a host-cell, by infecting or transforming a host  
 CC cell capable of expressing a chelon coding sequence with a vector  
 CC comprising a promoter active in the host cell operably linked to a coding  
 CC region for the protein to produce a recombinant host cell and culturing  
 CC the recombinant host cell under conditions, where DNA is expressed.  
 CC The nucleic acid encoding the chelon protein is useful for binding  
 CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
 CC metal ions from contaminated soil, ground water, hydroponic solutions or  
 CC irrigation water of waste streams. The DNA of the invention, when  
 CC immobilised onto a solid support, is useful for concentrating heavy metal  
 CC ions from contaminated environment waste streams or contaminated  
 CC aqueous medium including biological fluids. The nucleic acid, when  
 CC recombinantly expressed in enteric bacteria (which are nonpathogenic and  
 CC nonpathogenic), is suitable for use in the in vivo sequestration and  
 CC elimination of mercuric ion from gastrointestinal tracts of animals or  
 CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
 CC molecules of the invention are also useful in water treatment resins.  
 CC The nucleic acid of the invention is highly specific and binds divalent  
 CC cation such as mercury or cadmium with high affinity. The present amino  
 CC acid sequence represents one of a collection (AAU97553-AAU97560) of  
 CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
 CC This sequence is one of the heavy metal binding proteins termed chelons

CC of the invention.  
 XX  
 SQ Sequence 117 AA;  
 Query Match 98.4%; Score 603; DB 23; Length 117;  
 Best Local Similarity 98.3%; Pred. No. 3.1e-52;  
 Matches 115; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MTHCEASSLAHKLKDVREKMDLARMETVLSLVACAHARGNVSCPLIASIQSSGT 60  
 |||||  
 DB 1 MTHCEASSLAHKLKDVREKMDLARMETVLSLVACAHARGNVSCPLIASIQSSGT 60  
 |||||  
 QY 61 HCEASSLAHKLKDVREKMDLARMETVLSLVACAHARGNVSCPSAMSHPOFEK 117  
 |||||  
 DB 61 HCEASSLAHKLKDVREKMDLARMETVLSLVACAHARGNVSCPSAMSHPOFEK 117  
 |||||  
 RESULT 8  
 AAU97555  
 ID AAU97555 standard; Protein: 117 AA.  
 AC AAU97555;  
 DT 13-AUG-2002 (first entry)  
 DE Synthetic cadmium/mercury ion binding chelon protein #3.  
 KW Mercuric ion; contaminated soil; ground water; hydroponic solution;  
 KW irrigation water; waste stream; contaminated aqueous medium;  
 KW biological fluid; gastrointestinal tract; chelon protein;  
 KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
 KW heavy metal binding protein.  
 OS Synthetic.  
 PN MO200230962-A2.  
 PD 18-APR-2002.  
 PF 12-OCT-2001; 2001WO-US31819.  
 PR 12-OCT-2000; 2000US-240465P.  
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 PI Summers AO, Caguiat JJ;  
 PS MPI: 2002-435437/46.  
 PT Novel non-naturally occurring recombinant DNA molecule encoding a  
 chelon protein useful for binding divalent cation mercury from  
 contaminated soil, water, aqueous medium including biological fluids -  
 Claim 4; Page 22; 42pp; English.  
 CC The present invention relates to a new non-naturally occurring  
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
 CC which binds mercuric ions. The invention is useful for recombinantly  
 CC producing a protein in a host-cell, by infecting or transforming a host  
 CC cell capable of expressing a chelon coding sequence with a vector  
 CC comprising a promoter active in the host cell operably linked to a coding  
 CC region for the protein to produce a recombinant host cell and culturing  
 CC the recombinant host cell under conditions, where DNA is expressed.  
 CC The nucleic acid encoding the chelon protein is useful for binding  
 CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
 CC metal ions from contaminated soil, ground water, hydroponic solutions or  
 CC irrigation water of waste streams. The DNA of the invention, when  
 CC immobilised onto a solid support, is useful for concentrating heavy metal  
 CC ions from contaminated environment waste streams or contaminated  
 CC aqueous medium including biological fluids. The nucleic acid, when  
 CC recombinantly expressed in enteric bacteria (which are nonpathogenic and  
 CC nonpathogenic), is suitable for use in the in vivo sequestration and  
 CC elimination of mercuric ion from gastrointestinal tracts of animals or

CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AAU97553-AAU97560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.

XX Sequence 117 AA;

Query Match 98.0%; Score 601; DB 23; Length 117;  
Best Local Similarity 98.3%; Pred. No. 4.9e-52;  
Matches 115; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTHCEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPLIASLGSSGT 60  
DB 1 MTHCEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPLIASLGSSGT 60  
QY 61 HCEEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPSAMSHPOFEK 117  
DB 61 HCEEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPSAMSHPOFEK 117

## RESULT 9

AAU97554 standard; Protein; 118 AA.

AAU97554;

13-ADG-2002 (first entry)

Synthetic cadmium/mercury ion binding chelon protein #2.

Mercuric ion; contaminated soil; ground water; hydropionic solution;  
irrigation water; waste stream; contaminated aqueous medium;  
biological fluid; gastrointestinal tract; chelon protein;  
enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
heavy metal binding protein.

Synthetic.

WO200230962-A2.

18-APR-2002.

12-OCT-2001; 2001WO-US31819.

12-OCT-2000; 2000US-240465P.

(UYGE-) UNIV GEORGIA RES FOUND INC.

Summers AO, Cagulat JT;

WPI; 2002-435437/46.

Novel non-naturally occurring recombinant DNA molecule encoding a  
chelon protein useful for binding divalent cation mercury from  
contaminated soil, water, aqueous medium including biological fluids -

Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydropionic solutions or

CC Irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents one of a collection (AAU97553-AAU97560) of  
CC synthetic cadmium/mercury ion binding chelon proteins of the invention.  
CC This sequence is one of the heavy metal binding proteins termed chelons  
CC of the invention.

XX Sequence 118 AA;

Query Match 97.0%; Score 594.5; DB 23; Length 118;  
Best Local Similarity 97.5%; Pred. No. 2.2e-51;  
Matches 115; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 MTHCEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPLIASLGSSG 59  
DB 1 MTHCEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPLIASLGSSG 60  
QY 60 THCEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPSAMSHPOFEK 117  
DB 61 THCEASSLAHEKIKDVREKMDLARMETVLSLVACACHARKGNVSCPSAMSHPOFEK 118

## RESULT 10

AAU97551 standard; Protein; 144 AA.

AAU97551;

13-ADG-2002 (first entry)

Shigella flexneri wild-type Merr protein.

Mercuric ion; contaminated soil; ground water; hydropionic solution;  
irrigation water; waste stream; contaminated aqueous medium;  
biological fluid; gastrointestinal tract; chelon protein;  
enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
heavy metal binding protein; Merr.

Shigella flexneri.

WO200230962-A2.

18-APR-2002.

12-OCT-2001; 2001WO-US31819.

12-OCT-2000; 2000US-240465P.

(UYGE-) UNIV GEORGIA RES FOUND INC.

Summers AO, Cagulat JT;

WPI; 2002-435437/46.

Novel non-naturally occurring recombinant DNA molecule encoding a  
chelon protein useful for binding divalent cation mercury from  
contaminated soil, water, aqueous medium including biological fluids -

Disclosure; Page 20; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly

CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ion from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present amino  
CC acid sequence represents the Shigella flexneri wild-type MerR protein of  
CC the invention. This sequence was used in the methods of the invention for  
CC production of heavy metal binding proteins termed chelons.

SQ Sequence 144 AA;

Query Match 47.3%; Score 290; DB 23; Length 144;  
Best Local Similarity 96.6%; Pred. No. 3.8e-21;  
Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 THCEASSIAEHKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGSSG 59  
DB 80 THCEASSIAEHKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGGAG 137

RESULT 11

ID AAY01816 standard; protein; 144 AA.

XX AC AAY01816;

XX DT 30-JUN-1999 (first entry)

XX DE MerR protein which has selectivity towards heavy metal ions.

XX KM Heavy metal ion selectivity; metal ion-specific affinity sensor;

XX KM zinc; mercury; cadmium; copper; lead; environmental sample; medicine;  
XX food.

XX OS Pseudomonas aeruginosa.

XX PN WO914597-A1.

XX PD 25-MAR-1999.

XX PF 15-SEP-1998; 98WO-SE01638.

XX PR 15-SEP-1997; 97SE-0003315.

XX PA (BERG/) BERGGREN C.

XX PA (BONT/) BONTIDEAN I.

XX PA (CSOE/) CSOEREGT E.

XX PA (JOHA/) JOHANSSON G.

XX PA (MATI/) MATTIASSON B.

XX PA (UNBI/) UNIV BIRMINGHAM SCHOOL BIOLOGICAL SCI.

XX PA (VITO-) VITO VLAAMSE INSTELLING TECHNOLOGISCH.

XX PI Berggren C, Bontidean I, Brown N, Corbisier P, Csoeregi E;

XX PI Hobman J, Jakeman K, Johansson G, Lloyd J, Mattiasson B;

XX PI Van Der Lelle D, Wilsson J;

XX DR WPI; 1999-254424/21.

PT Capacitance sensor specific for heavy metal ions  
XX  
XX Claim 8; Page 22-23; 40pp; English.

CC The present sequence represents a protein that is selective towards  
CC heavy metal ions. It is used in the construction of the sensor of the  
CC invention. The specification describes a metal ion-specific, affinity  
CC sensor that measures capacitance. The sensor comprises a piece of  
CC noble metal to which are bound groups that bind specifically to selected  
CC heavy metal ions. These groups are bound to a self-assembling monolayer  
CC that covers at least 90%, more preferably at least 99% of the noble  
CC metal surface. The noble metal is a rod or piece of insulating material  
CC (glass, quartz or silica) on which a noble metal is sputtered. The  
CC sensor is used for qualitative or quantitative detection of selected  
CC heavy metal ions in liquid samples, particularly of zinc, mercury,  
CC cadmium, copper and lead in e.g. environmental samples, medicines,  
CC foods and other products.

SQ Sequence 144 AA;

Query Match 45.0%; Score 276; DB 20; Length 144;  
Best Local Similarity 93.0%; Pred. No. 9.2e-20;  
Matches 53; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 THCEASSIAEHKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGSS 58  
DB 80 THCEASSIAEHKLDVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGG 136

RESULT 12

ID AAR49668 standard; protein; 159 AA.

XX AC AAR49668;

XX DT 16-SEP-1994 (first entry)

XX DE Protein product of mercury resistance control gene merR(1).

XX KM Resistance; mercury; selectable marker; Thiobacillus ferrooxidans;

XX KM transformation; detection.

XX OS Thiobacillus ferrooxidans.

XX PN JP06000083-A.

XX PD 11-JAN-1994.

XX PF 17-JAN-1991; 91JP-0018338.

XX PR 17-JAN-1991; 91JP-0018338.

XX PA (AKIT-) AKITA KEN.

XX PA (DOWA) DOWA MINING CO LTD.

XX DR WPI; 1994-077131/10.

XX DR N-PSDB; AAQ58554.

XX PT Mercury resistant control gene merR and shuttle vector - for  
XX PT enhanced expression of mercury resistance marker in transformed  
XX PT Thiobacillus sp.

XX PS Disclosure; Page 2; 26pp; Japanese.

XX CC The mercury resistance genes can be used as selectable markers when  
XX CC used to transform other bacteria.

SQ Sequence 159 AA;

Query Match 41.9%; Score 257; DB 15; Length 159;  
Best Local Similarity 76.9%; Pred. No. 7.9e-18;  
Matches 50; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

OY 2 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQSSGTH 61  
DB 80 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQSSGTH 139  
OY 62 CEAS 66  
DB 140 GDRAT 144

## RESULT 13

AAR49669  
ID AAR49669 standard; Protein; 135 AA.  
XX AAR49669;  
AC AAR49669;  
DT 16-SEP-1994 (first entry)  
XX  
DE Protein product of mercury resistance control gene merR(2).  
XX  
KM Resistance; mercury; selectable marker; Thiobacillus ferrooxidans;  
KW transformation; detection.  
XX  
OS Thiobacillus ferrooxidans.  
XX  
PN JP06000083-A.  
XX  
PD 11-JAN-1994.  
XX  
PF 17-JAN-1991; 91JP-0018338.  
XX  
PR 17-JAN-1991; 91JP-0018338.  
XX  
PA (AKIT-) AKITA KEN.  
XX (DOMA ) DOMA MINING CO LTD.  
DR WPI; 1994-077131/10.  
DR N-PSDB; AAO58555.  
XX  
PT Mercury resistant control gene merR and shuttle vector - for  
PT enhanced expression of mercury resistance marker in transformed  
PT Thiobacillus sp.  
XX  
PS Disclosure; Page 3; 26pp; Japanese.  
XX  
CC The mercury resistance genes can be used as selectable markers when  
CC used to transform other bacteria.  
XX  
SQ Sequence 135 AA;

Query Match 20.4%; Score 125; DB 15; Length 135;  
Best Local Similarity 44.4%; Pred. No. 7.5e-05;  
Matches 24; Conservative 13; Mismatches 17; Indels 0; Gaps 0;

OY 2 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQ 55  
DB 76 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQ 129

## RESULT 14

AAR49670  
ID AAR49670 standard; Protein; 135 AA.  
XX AAR49670;  
AC AAR49670;  
DT 16-SEP-1994 (first entry)  
XX  
DE Protein product of mercury resistance control gene merR(3).  
XX  
KM Resistance; mercury; selectable marker; Thiobacillus ferrooxidans;  
KW transformation; detection.  
XX  
OS Thiobacillus ferrooxidans.  
XX

PN JP06000083-A.

XX 11-JAN-1994.

PD 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

OY 2 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQ 55  
DB 76 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQ 129

## RESULT 15

ABP38137  
ID ABP38137 standard; Protein; 138 AA.

XX ABP38137;

AC ABP38137;

DT 24-JUL-2002 (first entry)

XX

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:2982.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

Query Match 20.1%; Score 123; DB 15; Length 135;  
Best Local Similarity 44.4%; Pred. No. 0.00012;  
Matches 24; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

OY 2 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQ 55  
DB 76 THCEASSLAEHLKDVREKMDIARMEVYLSLVCAHARGNVCPLIASLQ 129

DISCLOSURE: SEQ ID 2982; 267pp; English.

ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
frame (ORF) nucleic acid sequences which encode the amino acid sequences  
given in ABP35124 to ABP37960. The S. epidermidis sequences have

CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.

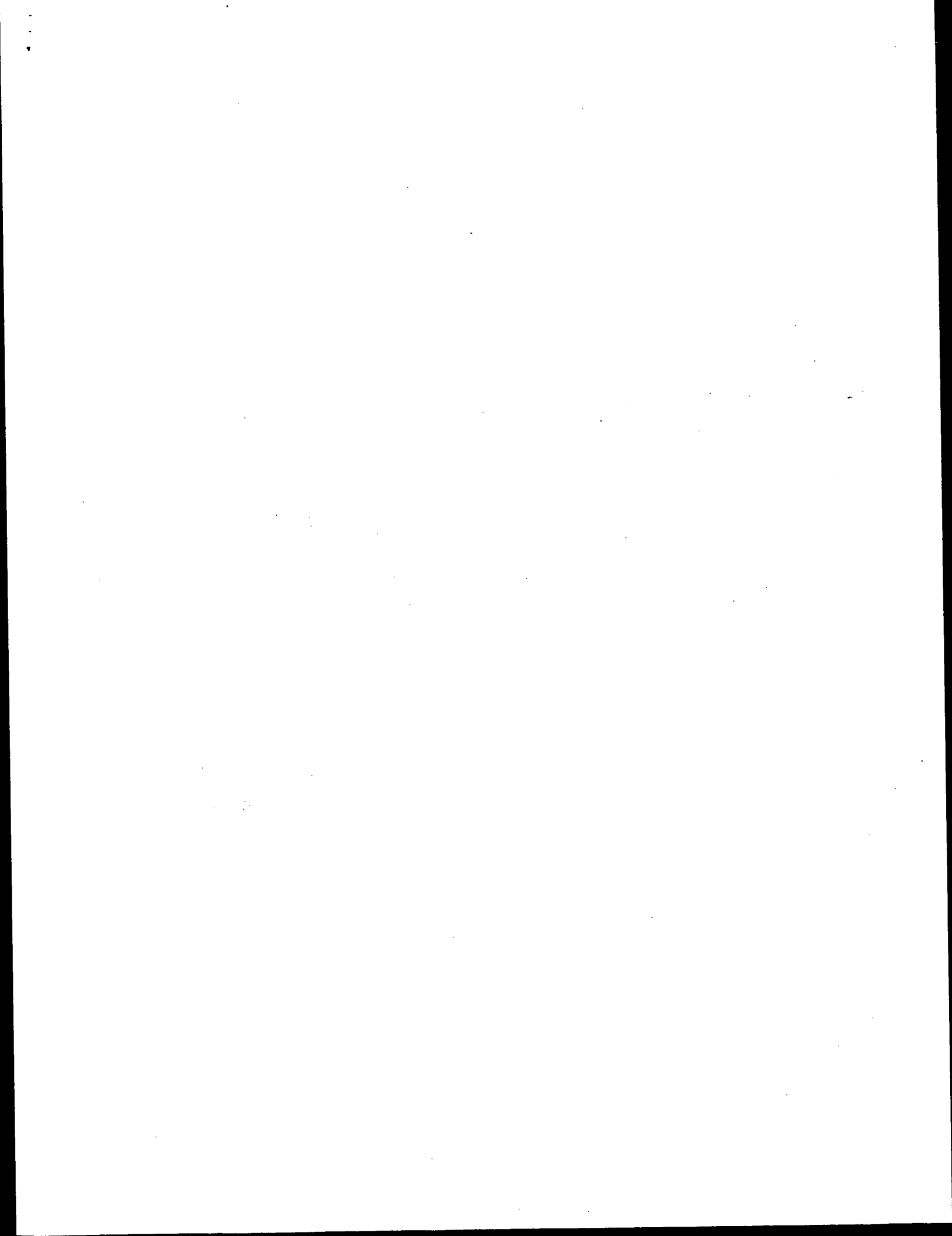
XX Sequence 138 AA;

Query Match 15.0%; Score 92; DB 23; Length 138;  
 Best Local Similarity 21.8%; Pred. No. 0.14;

Matches 27; Conservative 16; Mismatches 37; Indels 44; Gaps 2;

QY 28 METVLSLVAC-----HARKGNVSCP-----49  
 ||:|||||  
 Db 4 MEMKISELAKADVKNKETYRERKGLIAPRNSGYRIYSEETADRVRFIRKMKELDF 63  
 QY 50 -----LIASLOGSSGTHCEASSLAHRLKDVREKMDLARMETVLSLVACAHARKGN 103  
 ||:|||||  
 Db 64 SLKETIHLFGVVDGDERCKDMYAFVQKTEIERKVQDLRIQRLLELKEKCPDEKAI 123  
 QY 104 VSCP 107  
 :||  
 Db 124 YTCP 127

Search completed: May 28, 2003, 10:27:55  
 Job time : 37 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 28, 2003, 10:26:51 ; Search time 17 Seconds  
(without alignments)  
661.631 Million cell updates/sec

Title: US-09-977-137A-4

Perfect score: 613

Sequence: 1 MTCFEASSLAHEKLDVRE.....HARKGNVSCPSAMSHQPEK 117

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	290	47.3	144	2	S09527 regulatory protein
2	289	47.1	144	2	S51706 regulatory protein
3	280	45.7	144	2	S51720 regulatory protein
4	276	45.0	144	2	S51755 regulatory protein
5	276	45.0	144	2	S51749 regulatory protein
6	273	44.5	144	2	I39574 mer operon regulat
7	273	44.5	144	2	S32798 mer protein - Xan
8	273	44.5	144	2	S51756 regulatory protein
9	273	44.5	144	2	S37044 regulatory protein
10	269.5	44.0	151	2	S51703 regulatory protein
11	269.5	44.0	151	2	S51721 regulatory protein
12	269.5	44.0	151	2	S51705 regulatory protein
13	266.5	43.5	144	2	S51707 regulatory protein
14	260	42.4	144	2	A33858 mer protein - Esc
15	125	20.4	135	2	S18590 mer protein - Tn
16	123	20.1	135	2	S18584 mer protein - Tn
17	84	13.7	137	2	E97467 heavy metal depend
18	84	13.7	141	2	AH2685 transcription regu
19	83	13.5	135	2	A29504 hypothetical 16k p
20	83	13.5	141	2	AH1009 probable Zn(II)-re
21	81.5	13.3	1574	2	T13954 MEGF6 protein - ra
22	81	13.2	823	2	D86165 mercuric resistanc
23	79	12.9	132	2	A32239 mer protein F15K9.3 [i
24	79	12.9	132	2	T44501 mer protein F15K9.3 [i
25	79	12.9	141	2	A67892 zinc (II) responsi
26	79	12.9	141	2	E91148 zinc (II) responsi
27	79	12.9	141	2	A85994 probable transcrip
28	77	12.6	1727	2	T50073 myosin-like coiled
29	76.5	12.5	139	2	D82256 transcription regu

30	76.5	12.5	671	2	AF1294
31	76	12.4	674	2	T22733
32	75	12.2	141	2	AB0029
33	75	12.2	2442	2	T08621
34	73.5	12.0	671	2	AD1666
35	73	11.9	352	2	A39532
36	73	11.9	418	2	T40913
37	72.5	11.8	357	2	A71295
38	72.5	11.8	466	2	T29353
39	72.5	11.8	1694	2	T01896
40	72	11.7	197	2	T34551
41	72	11.7	626	2	E82314
42	72	11.7	790	2	T05576
43	72	11.7	2094	2	S31324
44	71.5	11.7	1238	2	T03465
45	71	11.6	132	2	T45509

## ALIGNMENTS

## RESULT 1

S09527 regulatory protein merR - plasmid NR1

C:Species: plasmid NR1

C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 20-Sep-1999

C:Accession: S09527

R:Barthelemy, P.; Gilbert, P.; Jackson, W.D.; Jones, C.S.; Summers, A.O.; Wisdom, S.

J. Mol. Appl. Genet. 2, 601-619, 1984

A:Title: The DNA sequence of the mercury resistance operon of the IncFII plasmid NR1.

A:Reference number: S07447; MUID:85159407; PMID:6530603

A:Accession: S09527

A:Molecule type: DNA

A:Residues: 1144 <BAR>

A:Cross-references: EMBL:K03089; NID:g150389; PIDN:AA859072.1; PID:g455296

C:Genetics:

A:Genome: plasmid

C:Superfamily: transcription repressor glrR

Query Match 47.3%; Score 290; DB 2; Length 144;

Best Local Similarity 96.6%; Pred. No. 7.2e-20;

Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQSGSG 59

DB 80 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQSGSG 137

## RESULT 2

S51706 regulatory protein merR - Alcaligenes faecalis

C:Species: Alcaligenes faecalis

C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999

C:Accession: S51706

R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.

submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from

A:Reference number: S51703

A:Accession: S51706

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1144 <OSB>

A:Cross-references: EMBL:Z333484; NID:g607038; PIDN:CA83892.1; PID:g607039

C:Superfamily: transcription repressor glrR

Query Match 47.1%; Score 289; DB 2; Length 144;

Best Local Similarity 96.6%; Pred. No. 8.9e-20;

Matches 56; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQSGSG 59

DB 80 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQSGSG 137

## RESULT 3

regulatory protein merR - Enterobacter aerogenes  
C:Species: Enterobacter aerogenes

C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
C:Accession: S51720

R:Osborn, A.M.; Bruce, K.D.; Striike, P.; Ritchie, D.A.

submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from me

A:Reference number: S51703

A:Accession: S51720

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <OSB>

A:Cross-references: EMBL:Z33488; NID:g607071; PIDN:CAA8386.1; PID:g607072

C:Superfamily: transcription repressor glnr

Query Match 45.7%; Score 280; DB 2; Length 144;  
Best Local Similarity 94.7%; Pred. No. 6e-19;  
Matches 54; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGSS 58  
|||||  
DB 80 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGGT 136

## RESULT 4

regulatory protein merR - Pseudomonas sp.

C:Species: Pseudomonas sp.

C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
C:Accession: S51755

R:Osborn, A.M.; Bruce, K.D.; Striike, P.; Ritchie, D.A.

submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from me

A:Reference number: S51703

A:Accession: S51755

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <OSB>

A:Cross-references: EMBL:Z33489; NID:g607167; PIDN:CAA83897.1; PID:g607168

C:Superfamily: transcription repressor glnr

Query Match 45.0%; Score 276; DB 2; Length 144;  
Best Local Similarity 93.0%; Pred. No. 1.4e-18;  
Matches 53; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGSS 58  
|||||  
DB 80 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGGA 136

## RESULT 5

regulatory protein merR - Pseudomonas fluorescens

C:Species: Pseudomonas fluorescens

C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 24-May-2001  
C:Accession: S51749

R:Osborn, A.M.; Bruce, K.D.; Striike, P.; Ritchie, D.A.

submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from me

A:Reference number: S51703

A:Accession: S51749

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <OSB>

A:Cross-references: EMBL:Z33490; NID:g607153; PIDN:CAA83898.1; PID:g607154

C:Superfamily: transcription repressor glnr

Query Match 45.0%; Score 276; DB 2; Length 144;  
Best Local Similarity 93.0%; Pred. No. 1.4e-18;  
Matches 53; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGSS 58  
|||||  
DB 80 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGGA 136

## RESULT 6

mer operon regulator - Alcaligenes sp.

C:Species: Alcaligenes sp.

C:Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Sep-1999  
C:Accession: I39574; S37035

R:Hobman, J.; Kholodil, G.; Nikiforov, V.; Ritchie, D.A.; Striike, P.; Yurleva, O.

Gene 146; 73-78, 1994

A:Title: The sequence of the mer operon of pMER327/419 and transposon ends of pMER327

A:Reference number: I39574; M01D:94341572; PMID:8063107

A:Accession: I39574

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-144 <RES>

A:Cross-references: EMBL:Z23094; GB:L20693; NID:g388553; PIDN:AA05979.1; PID:g388554

C:Genetics:

A:Gene: merR

C:Superfamily: transcription repressor glnr

Query Match 44.5%; Score 273; DB 2; Length 144;  
Best Local Similarity 82.8%; Pred. No. 2.7e-18;  
Matches 53; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGSSGTH 61  
|||||  
DB 80 THCEASSLAHKLKDYREKMDLARMETVISELVCAHARKGNVSCPLIASLOGKEPR 139

QY 62 CEBA 65  
-|  
DB 140 SADA 143

## RESULT 7

merR protein - Xanthomonas sp. transposon Tn5053

C:Species: Xanthomonas sp.

C:Date: 08-Dec-1993 #sequence\_revision 01-Dec-1995 #text\_change 20-Sep-1999  
C:Accession: S32798; S70142

R:Kholodil, G.Y.; Yurleva, O.V.; Lomovskaya, O.L.; Gorlenko, Z.M.; Mindlin, S.Z.; Nik

J. Mol. Biol. 230, 1103-1107, 1993

A:Title: Tn5053, a mercury resistance transposon with integron's ends.

A:Reference number: S32795; M01D:93253772; PMID:8387603

A:Accession: S32798

A:Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 1-144 <KHO>

A:Cross-references: EMBL:L03729; NID:g1019671; PIDN:AAA98396.1; PID:g154910

A:Experimental source: plasmid Rpl; transposon Tn5053

R:Kholodil, G.Y.; Mindlin, S.Z.; Bass, I.A.; Yurleva, O.V.; Minakhina, S.V.; Nikiforo

Mol. Microbiol. 17, 1189-1200, 1995

A:Title: Four genes, two ends, and a res region are involved in transposition of Tn50

A:Reference number: S70140; M01D:96130850; PMID:8594337

A:Accession: S70142

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-144 <KH2>

A:Cross-references: EMBL:L40585; NID:g710572; PIDN:AAA98322.1; PID:g710575

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995

C:Genetics:

A:Gene: merR

A:Mobile element: transposon Tn5053

C:Superfamily: transcription repressor glnr

Query Match 44.5%; Score 273; DB 2; Length 144;  
Best Local Similarity 82.8%; Pred. No. 2.7e-18;  
Matches 53; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY	2	THCEKSSLAERHTKQVREMDLAREYTLSELVYSCACHARKGVSCGLIASGSGTH	61
Db	80	THCEKSSLAERHTKQVREMDLAREYTLSELVYSCACHARKGVSCGLIASGSGTH	139
QY	62	CEEA	65
Db	140	SADA	143

**RESULT 8**

S51756  
regulatory protein merr - Comamonas testosteroni  
C:Species: Comamonas testosteroni  
C:Date: 07-May-1995 #sequence\_revision 10-Nov-1995 #text\_change 20-Sep-1999  
C:Accession: S51756  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
submitted to the EMBL Data Library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from me  
A:Reference number: S51703  
A:Accession: S51756  
A:Molecule type: DNA  
A:Residues: 1-144 <OSB>  
A:Cross-references: EMBL:Z33481; NID:g607169; PIDN:CAAB3889.1; PID:g607170  
A:Note: the source is given as Pseudomonas testosteroni  
C:Superfamily: transcription repressor glnR

Query Match	44.5%;	Score 273;	DB 2;	Length 144;
Best Local Similarity	82.8%;	Pred. NO. 2.7e-18;		
Matches 53; Conservative	3;	Mismatches 8;	Indels 0;	Gaps 0;

OY 2 THEEASSLAEHLKLDVREKADLARMETVLSEIWCACHRRKNVSCPLIASLGSSGTH 61  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 80 TICEASSLAELHKLTDVREMRADLARAEAVLSIDLCACCHARKNVSCPLIASLGKKRPR 139

QY	62	CEEA	65
		:1	
Db	140	SADA	143

## RESULT 9

S37044 regulatory protein merR - Pseudomonas fluorescens  
N:Alternate names: mer operon regulator  
C:Species: Pseudomonas fluorescens  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Sep-1999  
C:Accession: S37044  
R:Hobman, J., Kholodil, G., Nikiforov, V., Ritchie, D.A.; Strike, P.; Yurleva, O.  
submitted to the EMBL Data Library, June 1993  
A:Description: The nucleotide sequence of the mer operon of pmu100 and transposon ends  
A:Reference number: S37035  
A:Accession: S37044  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <HOB>  
A:Cross-references: EMBL:Z23095; NID:g397617; PIDN:CAAB0641.1; PID:g397618  
C:Superfamily: transcription repressor glnR

Query Match	44.5%	Score	273;	DB	2;	Length	144;
Best Local Similarity	82.8%	Pred. No.	2.7e-18;				
Matches	53;	Conservative	3;	Mismatches	8;	Indels	0;
						Gaps	0;

QY 2 THCEASSLAEHLKDVREKMDLARMETVISELVCACHARKGNVSCPLIASLGSSGTH 61  
|||||  
Db 80 THCEASSLAELHKLKDVREKMDLARMEAVISDLVCACHARKGNVSCPLIASLGKKRPR 139

QY	62	CEEA	65
		:1	
Db	140	SADA	143

RESULT 10

regulatory protein merr - *Acinetobacter calcoaceticus* (isolate SE11 and SE12)

C:Species: *Acinetobacter calcoaceticus*  
A:Variate: isolate SE11; isolate SE12  
C:date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999  
C:Accession: S51703; S51704  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from  
A:Reference number: S51703

A;Accession: S51703

A: Molecule type: DNA

```

..: molecule type: DNA
A:Residues: 1-151 <OSB>

```

A; Cross-references: EME

A; Experimental source: isolate SE11

A;Accession: S51704

A: Molecule type: DNA

```

molecule type: DNA
A:Residues: 1-151 <OSO>

```

**A;Cross-references: EMB**

A; Experimental source:

C;Genetics:

A;Gene: mekr

Superal Family: Transcription

Query Match

Best Local Similarity

Matches 57; Conser

QY 2 THCEASSLAHKK

80 THE EAST AFRICAN

00 LICENSING

QY 62 CFEASSLAEH 71

—  
—  
+ +  
—

D5 138 L-AASARGSH 14

## RESULT 11

S51721  
regulatory protein merr - Enterobacter cloacae  
C:Species: Enterobacter cloacae  
C:Date: 07-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Sep-1999  
C:Accession: S51721  
R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
submitted to the EMBL data library, May 1994  
A:Description: Sequence conservation between regulatory mercury resistance genes from  
A:Reference number: S51703  
A:Accession: S51721  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <OSB>  
A:Cross-references: EMBL:Z33486; NID:g607073; PIDN:CAA83894.1; PID:g607074  
C:Superfamily: transcrib. repressor glnR

Query Match	44.0%	Score 269.5	DB 2	Length 151
Best Local Similarity	81.4%	Pred. No. 5.9e-18		
Matches 57; Conservative	1	Mismatches 9	Indels 3	Gaps 2

QY 2 THCEASSIAEHKIKDVREKADLARHETVLSIVCACARKGNVSCPLIASLQSSCTH 61  
 |||||  
 Db 80 THCEASSIAEHKIKDVREKADLARHETVLSIVCACARKGNVSCPLIASLQ--DGTK 137

QY 62 CEASSLAEH 71  
||: |  
Db 138 L-AASARGSH 146

## RESULT 12

S51705 regulatory protein merr - Klebsiella oxytoca  
C:Species: Klebsiella oxytoca  
C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999  
Accession: S51705

R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.  
submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from me

A:Reference number: S51703

A:Accession: S51705

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <OSB>

A:Cross-references: EMBL:Z33485; NID:g607036; PIDN:CAA83893.1; PID:g607037

C:Superfamily: transcription repressor glr

Query Match 44.0%; Score 269.5; DB 2; Length 151;

Best Local Similarity 81.4%; Pred. No. 5.9e-18; Mismatches 9; Indels 3; Gaps 2;

Matches 57; Conservative 1; Mismatches 9; Indels 3; Gaps 2;

2 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLOGSSGTH 61

DB 80 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLQ--DGTG 137

QY 62 CEASSLAH 71

DB 138 L-AASARGSH 146

RESULT 13  
S51707  
regulatory protein merR - Agrobacterium radiobacter

C:Species: Agrobacterium radiobacter

C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 20-Sep-1999

C:Accession: S51707

R:Osborn, A.M.; Bruce, K.D.; Strike, P.; Ritchie, D.A.

submitted to the EMBL Data Library, May 1994

A:Description: Sequence conservation between regulatory mercury resistance genes from me

A:Reference number: S51703

A:Accession: S51707

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <OSB>

A:Cross-references: EMBL:Z33487; NID:g607040; PIDN:CAA83895.1; PID:g607041

C:Superfamily: transcription repressor glr

Query Match 43.5%; Score 266.5; DB 2; Length 151;

Best Local Similarity 80.0%; Pred. No. 1.1e-17; Mismatches 9; Indels 3; Gaps 2;

Matches 56; Conservative 2; Mismatches 9; Indels 3; Gaps 2;

QY 2 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLOGSSGTH 61

DB 80 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLQ--DGTG 137

QY 62 CEASSLAH 71

DB 138 L-AASARGSH 146

RESULT 14  
A33858  
merR protein - Escherichia coli plasmid pdu1358

C:Species: Escherichia coli

C:Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 20-Sep-1999

C:Accession: A33858

R:Nucifora, G.; Chu, L.; Silver, S.; Mitra, T.K.

J. Bacteriol. 171, 4241-4247, 1989

A:Title: Mercury operon regulation by the merR gene of the organomercurial resistance sy

A:Reference number: A33858; MUID:89327136; PMID:2666393

A:Accession: A33858

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-144 <NUC>

A:Cross-references: GB:M24940; NID:g150631; PIDN:AAA98221.1; PID:g455313

C:Genetics:

A:Genome: plasmid

C:Superfamily: transcription repressor glr

C:Keywords: DNA binding; transcription regulation

Query Match 42.4%; Score 260; DB 2; Length 144;

Best Local Similarity 92.7%; Pred. No. 4.2e-17; Mismatches 2; Indels 0; Gaps 0;

Matches 51; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLOG 56

DB 80 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLQ 134

RESULT 15

S18590  
merR protein - Thiobacillus ferrooxidans

C:Species: Thiobacillus ferrooxidans

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Sep-1999

C:Accession: S18590

R:Inoue, C.; Sugawara, K.; Kusano, T.

Mol. Microbiol. 5, 2707-2718, 1991

A:Title: The merR regulatory gene in Thiobacillus ferrooxidans is spaced apart from t

A:Reference number: S18584; MUID:92140035; PMID:1779760

A:Accession: S18590

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-135 <INO>

A:Cross-references: EMBL:X57326; NID:g48150; PIDN:CAA40603.1; PID:g48157

C:Superfamily: transcription repressor glr

Query Match 20.4%; Score 125; DB 2; Length 135;

Best Local Similarity 44.4%; Pred. No. 0.00012; Mismatches 17; Indels 0; Gaps 0;

Matches 24; Conservative 13; Mismatches 17; Indels 0; Gaps 0;

QY 2 THCEASSLAHKLKDYREKMDLARMETVSELVACACHARKGNVSCPLIASLQ 55

DB 76 THCEAAELASRHLASVOARLEHRIHALOKLEACNSOGNFSCPLIASLR 129

Search completed: May 28, 2003, 10:29:15  
Job time : 18 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:25:27 ; Search time 12 Seconds  
(without alignments)  
404.394 Million cell updates/sec

Title: US-09-977-137a-4  
Perfect score: 613  
Sequence: 1 MTHCEASLSAEHLKLDVRE.....HARKGNVSCPSAMSHPOFEK 117

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query length	ID	Description
1	290	47.3	144	MERR_SALTI
2	276	45.0	144	MERR_PSEAE
3	260	42.4	144	MERR_SERMA
4	83	13.5	135	MERR_STAU
5	79	12.9	132	MERR_BACSR
6	79	12.9	141	ZNTR_ECOTL
7	77	12.6	1727	ALML_SCHPO
8	76.5	12.5	139	CUER_VIBCH
9	75.5	12.3	799	ITBN_DROME
10	74.5	12.2	1070	Y352_HUMAN
11	73	11.9	352	WN3A_MOUSE
12	72.5	11.8	377	ALR_TREPA
13	72	11.7	245	R841_MOUSE
14	72	11.7	295	X769_HUMAN
15	72	11.7	2349	TFR_HUMAN
16	71.5	11.7	1238	SBQC_RHOCA
17	71	11.6	502	K2C8_XENLA
18	71	11.6	600	DNLI_ACIAM
19	71	11.6	1786	LMB1_MOUSE
20	71	11.6	3321	PCN2_HUMAN
21	70.5	11.5	129	HMMR_RHILV
22	70.5	11.5	770	STF3_HUMAN
23	70.5	11.5	770	STF3_MOUSE
24	70.5	11.5	770	STF3_MOUSE
25	70.5	11.5	1423	ALBU_PETMA
26	70	11.4	1102	MYSC_CHICK
27	70	11.4	3003	ZFH2_DROME
28	69.5	11.3	140	DHSB_ARATH
29	69	11.3	1935	MYH7_RAT
30	69	11.3	2547	PAFV_HUMAN
31	68.5	11.2	354	CD72_MOUSE
32	68.5	11.2	772	PMIP_YEAST
33	68.5	11.2	955	KINL_DEICH

34	68	11.1	244	1	MYH7_PAPHA	P11778	papio hamad
35	68	11.1	415	1	HEM1_METGO	Q92nd6	neisseria g
36	68	11.1	415	1	HEM1_METMA	Q92nd5	neisseria m
37	68	11.1	415	1	HEM1_METMB	P56994	neisseria m
38	68	11.1	833	1	IF2_PASMU	P57873	pasturella
39	68	11.1	1934	1	MYH7_MESAU	P13540	mesocricetu
40	68	11.1	1935	1	MYH7_HUMAN	P12883	homo sapien
41	68	11.1	1938	1	MYH6_RAT	P02363	homo sapien
42	67.5	11.0	195	1	PRD3_ARATH	P57741	arabidopsis
43	67.5	11.0	252	1	DHSB_SCHPO	P21911	schizosacch
44	67.5	11.0	513	1	TREPE_BACPU	P18267	bacillus pu
45	67	10.9	673	1	Y552_HUMAN	O60299	homo sapien

## ALIGNMENTS

```

RESULT 1
ID MERR_SALTI STANDARD: PRT: 144 AA.
AC P07044.
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mercuric resistance operon regulatory protein.
GN MERR OR HCM1.235
OS Salmonella typhi, and
OS Shigella flexneri.
OG Plasmid pHCM1, and plasmid IncFII NR1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CX Salmonella.
NCBI_TaxID=601, 623;
[1]
SEQUENCE FROM N.A.
SPECIES=S.typhi; STRAIN=CT18; PLASMID=pHCM1;
MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Baker S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Cronin A., Davis P., Brooks K., Chillingworth T., Connerton P.,
RA Feltwell T., Hamlin N., Haque R.M., Dowd L., White N., Farrar J.,
RA Krogg A., Larsen T.S., Leach S., Moulé S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
"Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
Nature 413:848-852(2001).
[2]
SEQUENCE FROM N.A.
SPECIES=S.flexneri; PLASMID=IncFII NR1; TRANSPONSON=Trn21;
MEDLINE=85159407; PubMed=6530603;
RA Barthelemy P., Gilbert P., Jackson W.J., Jones C.S., Summers A.O.,
RA Wisdom S.;
"The DNA sequence of the mercury resistance operon of the IncFII
plasmid NR1."
J. Mol. Appl. Genet. 2:601-619(1984).
-1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY
RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPRESSES
TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;
WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION
AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING
BOUND TO THE MER SITE.
-1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```

DR EMBL: AL513383; CAD09617.1; -  
 DR EMBL: K03089; AAB59072.1; -  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PR00040; HTHMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 DR Transcription regulation; Activator; Repressor; Mercuric resistance;  
 KM Mercury; DNA-binding; Plasmid; Transposable element;  
 KW Complete proteome.  
 FT DNA\_BIND 10 29 H-T-H MOTIF (POTENTIAL).  
 FT METAL 82 82 HG(2+).  
 FT METAL 117 117 HG(2+).  
 FT METAL 126 126 HG(2+).  
 SQ SEQUENCE 144 AA; 15905 MW; 8BEEC928A7B83EE9 CRC64;  
 Query Match 47.3%; Score 290; DB 1; Length 144;  
 Best Local Similarity 96.6%; Pred. No. 1.9e-20;  
 Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 TCCEASSLAEHKIKDYREKMDLARMETVLSLYCACHARGNVSCPLIASLOGSSG 59  
 DB 80 TCCEASSLAEHKIKDYREKMDLARMETVLSLYCACHARGNVSCPLIASLOGGAG 137  
 RESULT 2  
 MERR\_PSEAE STANDARD; PRT; 144 AA.  
 AC P06688;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Mercuric resistance operon regulatory protein.  
 GN Pseudomonas aeruginosa, and  
 OS Pseudomonas fluorescens.  
 OG Plasmid pVS1.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 NCBI\_TaxID=287, 294;  
 RX SEQUENCE FROM N.A.  
 RX MEDLINE=85014891; PubMed=6091128;  
 RX Misra T.K., Brown N.L., Fritlinger S., Pridmore R.D.,  
 Barnes W.M., Haberstroh L., Silver S.;  
 "Mercuric ion-resistance operons of plasmid R100 and transposon  
 Tn501: the beginning of the operon including the regulatory region  
 and the first two structural genes.";  
 Proc. Natl. Acad. Sci. U.S.A. 81:5975-5979(1984).  
 [12]  
 RX SEQUENCE FROM N.A.  
 RX STRAIN=SB4;  
 RX Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
 Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 [13]  
 RX MEDLINE=90001158; PubMed=2551364;  
 RX Shewchuk L.M., Verdine G.L., Nash H., Walsh C.T.;  
 "Mutagenesis of the cysteines in the metalloregulatory protein Merr  
 indicates that a metal-bridged dimer activates transcription.";  
 Biochemistry 28:6140-6145(1989).  
 -1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
 RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPASSES  
 TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
 WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
 AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
 BOUND TO THE MER SITE.  
 -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: Z00027; CA07320.1; -  
 DR EMBL: Z33489; CA083897.1; -  
 DR EMBL: Z33490; CA083898.1; -  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PR00040; HTHMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 DR Transcription regulation; Activator; Repressor; Mercuric resistance;  
 KM Mercury; DNA-binding; Plasmid; Transposable element.  
 KW Mercury; DNA-binding; Plasmid; Transposable element.  
 FT DNA\_BIND 10 29 H-T-H MOTIF (POTENTIAL).  
 FT METAL 82 82 HG(2+).  
 FT METAL 117 117 HG(2+).  
 FT METAL 126 126 HG(2+).  
 FT MUTAGEN 82 82 C->A: ABOLISHES TRANSCRIPTIONAL  
 FT MUTAGEN 115 115 ACTIVATION.  
 FT MUTAGEN 117 117 C->A: SLIGHT INCREASE IN TRANSCRIPTIONAL  
 FT MUTAGEN 117 117 ACTIVATION.  
 FT MUTAGEN 126 126 C->A: DECREASE IN TRANSCRIPTIONAL  
 FT MUTAGEN 126 126 ACTIVATION.  
 FT MUTAGEN 126 126 C->S: ABOLISHES TRANSCRIPTIONAL  
 FT SEQUENCE 144 AA; 15763 MW; C573298A8FF0846EF CRC64;  
 Query Match 45.0%; Score 276; DB 1; Length 144;  
 Best Local Similarity 93.0%; Pred. No. 3.9e-19;  
 Matches 53; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 TCCEASSLAEHKIKDYREKMDLARMETVLSLYCACHARGNVSCPLIASLOGSS 58  
 DB 80 TCCEASSLAEHKIKDYREKMDLARMETVLSLYCACHARGNVSCPLIASLOGGA 136  
 RESULT 3  
 MERR\_SERMA STANDARD; PRT; 144 AA.  
 AC P13111;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE Mercuric resistance operon regulatory protein.  
 GN Serratia marcescens.  
 OS Serratia marcescens.  
 OG Plasmid pD01358.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Serratia.  
 NCBI\_TaxID=615;  
 RX SEQUENCE FROM N.A.  
 RX MEDLINE=89327136; PubMed=2666393;  
 RX Nucifora G., Chu L., Silver S., Misra T.K.;  
 "Mercury operon regulation by the merr gene of the organomercurial  
 resistance system of plasmid pD01358.";  
 J. Bacteriol. 171:4241-4247(1989).  
 -1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
 RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPASSES  
 TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
 WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
 AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
 BOUND TO THE MER SITE.  
 -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M24940; AAA98221.1; -  
DR PIR: A33858; A33858.  
DR InterPro: IPR000551; HTH\_MeRR.  
DR Pfam: PF00376; merr; 1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR; 1.  
KW Transcription regulation; Activator; Repressor; Mercuric resistance;  
KW Mercury; DNA-binding; Plasmid.  
FT DNA\_BIND 10 29 H-T-H MOTIF (POTENTIAL).  
FT METAL 82 82 HG(2+).  
FT METAL 117 117 HG(2+).  
FT METAL 126 126 HG(2+).  
SQ SEQUENCE 144 AA; 16033 MW; 05FBF5224B89C052 CRC64;

Query Match 42.4%; Score 260; DB 1; Length 144;  
Best Local Similarity 92.7%; Pred. No. 1, 2e-17;  
Matches 51; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 THCEASSIAEHKLDVREKMDLARETVLSELYCACHARKGNVSCPLIASLOG 56  
DB 80 THCEASSIAEHKLDVREKMDLARETVLSELYCACHARKGNVSCPLIASLOG 134

RESULT 4  
MERR\_STAAU STANDARD; PRT; 135 AA.  
ID MERR\_STAAU P22874;  
AC P22874;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE Mercuric resistance operon regulatory protein.  
GN MERR.  
OS Staphylococcus aureus.  
OG Plasmid p1258.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87260937; PubMed=3037534;  
RA Laddaga R.A., Chu L., Misra T.K., Silver S.;  
RT "Nucleotide sequence and expression of the mercurial-resistance  
RT operon from Staphylococcus aureus plasmid p1258.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:5106-5110(1987).  
CC -1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
CC RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPASSES  
CC TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
CC WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
CC AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
CC BOUND TO THE MER SITE.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L29436; AAA98241.1; -  
DR PIR: A29504; A29504.  
DR InterPro: IPR000551; HTH\_MeRR.  
DR Pfam: PF00376; merr; 1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR; 1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.

KW Transcription regulation; Activator; Repressor; Mercuric resistance;  
KW Mercury; DNA-binding; Plasmid.  
FT DNA\_BIND 5 24 H-T-H MOTIF (POTENTIAL).  
FT METAL 79 79 HG(2+).  
FT METAL 114 114 HG(2+).  
FT METAL 123 123 HG(2+).  
SQ SEQUENCE 135 AA; 15741 MW; FDC1A852621D4F82 CRC64;

Query Match 13.5%; Score 83; DB 1; Length 135;  
Best Local Similarity 30.2%; Pred. No. 0.29;  
Matches 16; Conservative 12; Mismatches 25; Indels 0; Gaps 0;

OY 4 CEASSIAEHKIKDVKREKMDLARETVLSELYCACHARKGNVSCPLIASLOG 56  
DB 79 CDMYAFYQKTEIKRKVQGLRIQRLBELKEKCPDEKAMYTPTITLMG 131

RESULT 5  
MERR\_BACSR STANDARD; PRT; 132 AA.  
ID MERR\_BACSR P22853;  
AC P22853;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE Mercuric resistance operon regulatory protein.  
GN MERR.  
OS Bacillus sp. (strain RC607).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89123092; PubMed=2536669;  
RA Wang Y., Moore M., Levinson H.S., Silver S., Walsh C., Mahler I.;  
RT "Nucleotide sequence of a chromosomal mercury resistance determinant  
RT from a Bacillus sp. with broad-spectrum mercury resistance.";  
RL J. Bacteriol. 171:83-92(1989).  
RN [2]  
RP FUNCTION.  
RX MEDLINE=89123021; PubMed=2492496;  
RA Helmann J.D., Wang Y., Mahler I., Walsh C.T.;  
RT "Homologous metalloregulatory proteins from both Gram-positive and  
RT Gram-negative bacteria control transcription of mercury resistance  
RT operons.";  
RL J. Bacteriol. 171:222-229(1989).  
RN [3]  
RP MUTAGENESIS.  
RX MEDLINE=90161989; PubMed=2305262;  
RA Helmann J.D., Ballard B.T., Walsh C.T.;  
RT "The Mer metalloregulatory protein binds mercuric ion as a  
RT tricoordinate, metal-bridged dimer.";  
RL Science 247:946-948(1990).  
CC -1- FUNCTION: MEDIATES THE MERCURIC-DEPENDENT INDUCTION OF MERCURY  
CC RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPASSES  
CC TRANSCRIPTION BY BINDING TIGHTLY TO THE MER OPERATOR REGION;  
CC WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION  
CC AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING  
CC BOUND TO THE MER SITE.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF138877; AAA83973.1; -  
DR PIR: A32227; A32227.  
DR PIR: A32239; A32239.  
DR InterPro: IPR000551; HTH\_MeRR.







DB 84 AOEKOEISRKLEMTKOELEWIASCPGDSG--DCPITEQK--HC 130

RESULT 9

ID ITBN\_DROME STANDARD: PRT: 799 AA.

AC Q27591; Q9VIG7; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Integrin beta nu precursor.

GN BETA-INT-NU OR CG1762.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

SEQUENCE FROM N.A.

RC TISSUE-Midgut endoderm;

RX MEDLINE-94357079; PubMed=8076521;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chamee C.R., Pfeiffer B.D.,

RA Wan H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abail J.F., Agbayani A., An H.-J., Andrews-Plamkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,

RA Borkins C.D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fostler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,

RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Slier E., Spadling A.C., Stapleton M., Strong R., Sun E.,

RA Syrakas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RL The genome sequence of Drosophila melanogaster.

SC Science 287:2185-2195(2000).

CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN CELL ADHESION.

CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: L13305; AAC37169.1; -

DR EMBL: AE003669; AAF53952.1; -

DR HSSP: P05106; IJY2.

DR Flybase: FBgn0010395; beta-Int-nu.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR002369; Integrin\_B.

DR InterPro: IPR001169; Integrin\_beta\_C.

DR InterPro: IPR002049; Laminin\_EGF.

DR InterPro: IPR003659; Plectin-like.

DR InterPro: IPR002035; VWF\_A.

DR Pfam: PF00362; Integrin\_B: 1.

DR PRINTS: PR00011; EGF\_LAMININ.

DR PRODOM: PD001811; Integrin\_B: 1.

DR SMART: SM00181; EGF: 1.

DR SMART: SM00187; INB: 1.

DR SMART: SM00423; PSI: 1.

DR SMART: SM00327; VWFA: 1.

DR PROSITE: PS000022; EGF\_1; UNKNOWN\_4.

DR PROSITE: PS01186; EGF\_2; UNKNOWN\_4.

DR PROSITE: PS00243; INTEGRIN\_BETA: 1.

DR PROSITE: PS50234; VWFA: 1.

KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;

KW Signal.

FT SIGNAL 1 26

FT CHAIN 27 799

FT DOMAIN 726 725

FT TRANSMEM 747 746

FT DOMAIN 136 372

FT CARBOHYD 73 73

FT CARBOHYD 167 409

FT CARBOHYD 409 409

FT CARBOHYD 505 505

FT CARBOHYD 505 505

FT CARBOHYD 655 655

FT CONFLICT 680 680

FT CONFLICT 701 701

FT SEQUENCE 799 AA; 90841 MW; 351699D523F07DEB CRC64;

Query Match 12.3%; Score 75.5; DB 1; Length 799;

Best Local Similarity 26.1%; Pred. No. 9.5;

Matches 30; Conservative 15; Mismatches 35; Indels 35; Gaps 7;

QY 2 TRCE---EASSLAHK--LKDVEKMDLARMETVLSVYACAHKGNVSGPLIASIQG 56

DB 495 TTCNCTDNTNVTNSNELLQCKRPPSDKS-----TSELVCSNH---GDCDCGTCLDPG 546

QY 57 SSGTCEASSLAEHLKDKVERKMDLARMETVLSVYACAHKGNVSGPLIASIQG 111

DB 547 YNGPFECEHCL-----DDEKDLA-----CEC-----GCCVCAKYGWS 579

RESULT 10

ID Y355\_HUMAN STANDARD: PRT: 1070 AA.

AC O15063;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein KIAA0355.

GN KIAA0355.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

```

RP SEQUENCE FROM N.A.
RC [1]
RC TISSUE=Brain;
RA MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/isb-sib.ch).
CC or send an email to license@isb-sib.ch).
CC -----
CC CC
CC DR EMBL; AB002353; BAA20812.1; -
CC KW Hypothetical protein.
CC FT DOMAIN 551 555 POLY-SER.
CC FT DOMAIN 695 700 POLY-PRO.
CC FT DOMAIN 1037 1044 POLY-PRO.
CC SO SEQUENCE 1070 AA; 116047 MW; 2D35C127C5EBA2F7 CRC64;

Query Match 12.28; Score 74.5; DB 1; Length 1070;
Best Local Similarity 22.08; Pred. No. 16;
Matches 33; Conservative 22; Mismatches 48; Indels 47; Gaps 7;

QY 4 CEEASLAEHLTKVREKMDLARETYLSELYCAC-HARKGNVSC---PLIASLOGSSG 59
QY |:::::||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 259 CSQSAIPHEQLKELNLTIDSLAQYKIALSLGHCYAKMGFHLNPKALIASLOGC-- 316

QY 60 TRCEAAS-----LAEHKLDV--REKMDLARET-VLSEL 93
QY |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 317 --CSEAEAGQGTGRGTPPQPMQCELPYTPVYQIGSHFLKGVSTNEASADNLKKTHTMQL 374

QY 94 VCACHARKG-----NVSCPSAW 110
QY |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 375 MKEAGCYNGITSRDDEPYTEVLNVCPSPTW 404

RESULT 11
WN3A_MOUSE STANDARD; PRT; 352 AA.
AC P27467;
AC 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WNT-3A protein precursor.
DE WNT3A OR WNT-3A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RC MEDLINE=91160971; PubMed=2001840;
RA Roelink H., Nusse R.;
RA "Expression of two members of the Wnt family during mouse
RA development -- restricted temporal and spatial patterns in the
RA developing neural tube.";
RA Genes Dev. 5:381-388(1991).
RL
CC -I- FUNCTION: LIGAND FOR MEMBERS OF THE FRIZZLED FAMILY OF SEVEN
CC TRANSMEMBRANE RECEPTORS. WNT-3 AND WNT-3A PLAY DISTINCT ROLES IN
CC CELL-CELL SIGNALING DURING MORPHOGENESIS OF THE DEVELOPING NEURAL
CC TUBE.
CC -I- SUBCELLULAR LOCATION: Possibly secreted and associates with the
CC extracellular matrix.
CC -I- TISSUE SPECIFICITY: DORSAL PORTION OF THE NEURAL TUBE (DEVELOPING
CC ROOF PLATE), AND MESENCHYME TISSUE SURROUNDING THE UMBILICAL

```

[illegible]



QY 4 CEASSLAEHLKDYREKMDLARMETVLSLYCACHARK-GNVC---PL-IASLOGSS 58  
 DB 186 CSSCSLSMHE-PEDTREFK-----NEALGELSSPHLOLMEPCDPWPLDMQPLLNKQ 238  
 QY 59 GTHCEASSLAEHLKDYREKMDLARMETVLSLYCACHA 99  
 DB 239 SDDQWQASASAK---SEEEKLAEIARQ---LOESAKLIA 273

RESULT 15  
 TPR\_HUMAN STANDARD: PRT: 2349 AA.  
 ID TPR\_HUMAN P12270;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nucleoprotein TPR.  
 GN TPR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93064711; PubMed=1437155;  
 RA Mitchell P.J., Cooper C.S.;  
 RT "The human tpr gene encodes a protein of 2094 amino acids that has  
 extensive coiled-coil regions and an acidic C-terminal domain.";  
 RT Oncogene 7:2329-2333(1992).  
 RL [2]  
 RP REVISIONS, AND CHARACTERIZATION.  
 RX MEDLINE=95096166; PubMed=7798308;  
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,  
 RA Sapphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;  
 RT "Tpr, a large coiled coil protein whose amino terminus is involved in  
 activation of oncogenic kinases, is localized to the cytoplasmic  
 surface of the nuclear pore complex.";  
 RT J. Cell Biol. 127:1515-1526(1994).  
 RL [3]  
 RP SEQUENCE OF 1-142 FROM N.A.  
 RX MEDLINE=88262257; PubMed=3387099;  
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;  
 RT "Tpr homologues activate met and raf.";  
 RT Oncogene 2:617-619(1988).  
 RL [4]  
 CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE  
 CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS  
 CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE  
 CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH  
 CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER  
 CC COMPONENTS, INCLUDING P62.  
 CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND  
 CC BRAIN. LOWER LEVELS IN HEART, LIVER, AND KIDNEY.  
 CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK  
 CC OR RAF GENES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X66397; CAA47021.1; -;  
 CC EMBL: Y00672; CAA68681.1; -;  
 CC PIR: S00928; S00928.  
 CC DR GeneW; HGNC:12017; TPR.  
 CC MIM: 189940; -;  
 CC KW Coiled coil; Proto-oncogene; Chromosomal translocation;  
 CC Nuclear protein; Transport.  
 CC FT DOMAIN 78 360 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 422 571 COILED COIL (POTENTIAL).

FT DOMAIN 575 628 COILED COIL (POTENTIAL).  
 FT DOMAIN 758 805 COILED COIL (POTENTIAL).  
 FT DOMAIN 834 869 COILED COIL (POTENTIAL).  
 FT DOMAIN 934 979 COILED COIL (POTENTIAL).  
 FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).  
 FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).  
 FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).  
 FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).  
 FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).  
 FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).  
 FT DOMAIN 527 530 POLY-SER.  
 FT DOMAIN 1833 1836 POLY-GLU.  
 FT DOMAIN 1957 1964 POLY-ASP.  
 FT DOMAIN 2295 2298 POLY-SER.  
 SQ SEQUENCE 2349 AA; 265600 MW; AFD0685CEDCA99F CRC64;

Query Match 11.7%; Score 72; DB 1; Length 2349;  
 Best Local Similarity 28.4%; Pred. No. 63;  
 Matches 33; Conservative 19; Mismatches 44; Indels 20; Gaps 6;

QY 4 CEASSLA-EHLKDYREKMDLARMETVLSLYCACHARKGNVSCPLIASLOGSSGTHC 62  
 DB 1068 COEQAKIAVEAONKYERELMLHAADVE-----ALQAKAEQVS--KMASVR---QHL 1113

QY 63 BEASSLAEHLKDYREKMDLARM-ETVLSLYCACHARKGNVSCPSAMSHPOFEK 117  
 DB 1114 EETTKAESQLIECAKSWERERMLKDEYKVCRCR-----DLEKONRLHDIK 1165

Search completed: May 28, 2003, 10:28:14  
 Job time : 14 secs





DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Regulatory protein.  
OS Alcaligenes faecalis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Alcaligenes.  
OX NCBI\_Taxid=511;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SE20;  
ID Osborn A.M., Bruce K.D., Ritchie D.A.;  
RA "Sequence Conservation between Regulatory Mercury Resistance Genes  
RT from Mercury Polluted and Pristine Environments."  
RL Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC EMBL: Z33484; CAAB3892.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr. 1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR. 1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY. 1.  
KW DNA-binding; Transcription regulation.  
SQ SEQUENCE 144 AA; 15832 MW; B71D7EF3C980DC49 CRC64;  
Query Match 47.1%; Score 289; DB 2; Length 144;  
Best Local Similarity 96.6%; Pred. No. 7.4e-21;  
Matches 56; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 THCEASSLAHKLQVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGSSG 59  
DB 80 THCEASSLAHKLQVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGSSG 137  
ID 046655 PRELIMINARY; PRT; 144 AA.  
AC 046655;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Regulatory protein.  
GN MERR.  
OS Enterobacter aerogenes (Aerobacter aerogenes).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Enterobacter.  
OX NCBI\_Taxid=548;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T238;  
ID Osborn A.M., Bruce K.D., Ritchie P., Ritchie D.A.;  
RA "Sequence Conservation between Regulatory Mercury Resistance Genes  
RT from Mercury Polluted and Pristine Environments."  
RL Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC EMBL: Z33488; CAAB3896.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr. 1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR. 1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY. 1.  
KW DNA-binding; Transcription regulation.  
SQ SEQUENCE 144 AA; 15652 MW; 1DE61F50D37A1337 CRC64;  
Query Match 45.7%; Score 280; DB 2; Length 144;  
Best Local Similarity 94.7%; Pred. No. 5.6e-20;  
Matches 54; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 2 THCEASSLAHKLQVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGSS 58  
ID 046655 PRELIMINARY; PRT; 144 AA.  
AC 046655;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Regulatory protein.  
GN MERR.  
OS Enterobacter aerogenes (Aerobacter aerogenes).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Enterobacter.  
OX NCBI\_Taxid=548;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T238;  
ID Osborn A.M., Bruce K.D., Ritchie P., Ritchie D.A.;  
RA "Sequence Conservation between Regulatory Mercury Resistance Genes  
RT from Mercury Polluted and Pristine Environments."  
RL Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC EMBL: Z33488; CAAB3896.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr. 1.  
DR PRINTS: PRO0040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR. 1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY. 1.  
KW DNA-binding; Transcription regulation.  
SQ SEQUENCE 144 AA; 15652 MW; 1DE61F50D37A1337 CRC64;

DB 80 THCEASSLAHKLQVREKMDLARMETVLSLVCAHARKGNVSCPLIASLOGCT 136  
RESULT 4  
ID 052395 PRELIMINARY; PRT; 144 AA.  
AC 052395;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Putative Merr protein (Organomercurial resistance regulatory  
protein).  
GN MERR.  
OS Pseudomonas putida, and  
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
OC Plasmid pMW0, and Plasmid pPB.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_Taxid=303, 316;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.putida; PLASMID=PMW0;  
RA "Greated A. Lamberton L., Williams P.A., Thomas C.M.;  
RT "Complete nucleotide sequence of IncP-9 plasmid pMW0."  
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.putida; STRAIN=OX; PLASMID=PPB;  
RX MEDLINE=85014891; PubMed=6091128;  
RA Mista T.K., Brown N.L., Fritzinger D.C., Primmore R.D., Barnes W.M.,  
RA Haberstroh L., Silver S.;  
RT "Mercuric ion-resistance operons of plasmid R100 and transposon Tn501:  
RT the beginning of the operon including the regulatory region and the  
RT first two structural genes."  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5975-5979(1984).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;  
RX MEDLINE=86174347; PubMed=3007931;  
RA Brown N.L., Mista T.K., Minnie J.N., Schmidt A., Self M., Silver S.;  
RT "The nucleotide sequence of the mercuric resistance operons of plasmid  
RT R100 and transposon Tn501: further evidence for mer genes which  
RT enhance the activity of the mercuric ion detoxification system."  
RL Mol. Gen. Genet. 202:143-151(1986).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;  
RX MEDLINE=93253772; PubMed=8387603;  
RA Khododli G.Ya., Yurleva O.V., Lomovskaya O.L., Gorienko Zh.M.,  
RA Mindlin S.Z., Nikiforov V.G.;  
RT "Tn5053, a mercury resistance transposon with integrin's ends."  
RL J. Mol. Biol. 230:1103-1107(1993).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;  
RX MEDLINE=96105204; PubMed=8529897;  
RA Reniero D., Galli E., Barbieri P.;  
RT "Cloning and comparison of mercury- and organomercurial-resistance  
RT determinants from a Pseudomonas stutzeri plasmid."  
RL Gene 166:77-82(1995).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.stutzeri; STRAIN=OX; PLASMID=PPB;  
RX MEDLINE=98148002; PubMed=9479042;  
RA Reniero D., Mozzon E., Galli E., Barbieri P.;  
RT "Two aberrant mercury resistance transposons in the Pseudomonas  
RT stutzeri plasmid pPB."  
RL Gene 208:37-42(1998).  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC EMBL: AJ344068; CAC86844.1; -;  
DR EMBL: U90263; AAC38229.1; -;  
DR InterPro: IPR000551; HTH\_Merr.



DR Pfam: PF00376; merr: 1.  
DR PRINTS: PR00040; HTMERR.  
DR SMART: SM00422; HTH\_MERR: 1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY: 1.  
KW Plasmid: DNA-binding; Transcription regulation.  
SQ SEQUENCE 144 AA; 15884 MW; F5760BEC8602RC7 CRC64;

Query Match 45.0%; Score 276; DB 2; Length 144;  
Best Local Similarity 84.4%; Pred. No. 1.4e-19;  
Matches 54; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 2 THCEASSTLAERKLDVREKMDLARETVLSLVCACHARKGNVSCPLIASLOGSSGTH 61  
DB 80 THCEASSTLAERKLDVREKMDLARETVLSLVCACHARKGNVSCPLIASLOGKKRPR 139  
OY 62 CEEA 65  
DB 140 SADA 143

## RESULT 5

O9AFK4 PRELIMINARY; PRT; 172 AA.  
AC O9AFK4;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Tn501 repressor.  
GN MERR.  
OS Shigella flexneri.  
OC Plasmid virulence plasmid pWR501.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21189246; PubMed=11292750;  
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grothbeck E.J., Burland V.,  
RA Plattner F.R.;  
RT "Complete DNA sequence and analysis of the large virulence plasmid of  
RT Shigella flexneri.";  
RL Infect. Immun. 69:3271-3285(2001).  
CC -!- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL: AF348706; AAK18578.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr: 1.  
DR PRINTS: PR00040; HTMERR.  
DR SMART: SM00422; HTH\_MERR: 1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY: 1.  
KW DNA-binding; Plasmid; Transcription regulation.  
SQ SEQUENCE 172 AA; 18826 MW; 897D139E7BC182A9 CRC64;

Query Match 45.0%; Score 276; DB 2; Length 172;  
Best Local Similarity 93.0%; Pred. No. 1.6e-19;  
Matches 53; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 THCEASSTLAERKLDVREKMDLARETVLSLVCACHARKGNVSCPLIASLOGSS 58  
DB 108 THCEASSTLAERKLDVREKMDLARETVLSLVCACHARKGNVSCPLIASLOGA 164

RESULT 6

O9R9W9 PRELIMINARY; PRT; 110 AA.  
AC O9R9W9;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE MERR (Fragment).  
GN Pseudomonas putida.  
OS Plasmid group 5 plasmid.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KT2440;  
RA Bruce K.D., Lilley A.K., Bailey M.J.;  
RT "mer sequences on plasmids";  
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF092070; AAD52706.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR SMART: SM00422; HTH\_MERR: 1.  
KW Plasmid.  
FT NON\_TER 1 1  
FT NON\_TER 110 110  
SQ SEQUENCE 110 AA; 12211 MW; 2C7C09EB8ACB7BCA CRC64;

Query Match 44.5%; Score 273; DB 2; Length 110;  
Best Local Similarity 82.8%; Pred. No. 2e-19;  
Matches 53; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 2 THCEASSTLAERKLDVREKMDLARETVLSLVCACHARKGNVSCPLIASLOGSSGTH 61  
DB 47 THCEASSTLAERKLDVREKMDLARETVLSLVCACHARKGNVSCPLIASLOGKKRPR 106  
OY 62 CEEA 65  
DB 107 SADA 110

## RESULT 7

O60233 PRELIMINARY; PRT; 144 AA.  
AC O60233; P75015; Q4314; Q51768; Q51808; Q52601; Q56390; Q56443;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Mercuric resistance operon regulatory protein (MERR protein).  
GN MERR.  
OS Alcaligenes sp.,  
OS Pseudomonas fluorescens,  
OS Comamonas testosteroni (Pseudomonas testosteroni),  
OS unidentified,  
OS Xanthomonas, and  
OS Pseudomonas sp. (strain ADP).  
OC Plasmid pMR327, Plasmid Rpl, and Plasmid pADP-1.  
OC Bacteria; Proteobacteria; Delta subdivision; Alcaligenaceae;  
OC Alcaligenes.  
OX NCBI\_TaxID=512, 294, 285, 32644, 338, 47660;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Alcaligenes sp., and P. fluorescens; TRANSPOSON=TN5053;  
RX MEDLINE=94341572; PubMed=8063107;  
RA Hobman J., Kholodil G., Nikiforov V., Ritchie D.A., Strike P.,  
RA Yurleva O.;  
RT "The sequence of the mer operon of pMR327/419 and transposon ends of  
RT pMR327/419, 330 and 05.";  
RL Gene 146:73-78(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=C. testosteroni; STRAIN=SE3;  
RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=unidentified;  
RX MEDLINE=96130850; PubMed=8594337;  
RA Kholodil G.Y., Mindlin S.Z., Bass I.A., Yurleva O.V., Minakhina S.V.,  
RA Nikiforov V.G.;  
RT "Four genes, two ends, and a res region are involved in transposition  
RT of Tn5053: a paradigm for a novel family of transposons carrying  
RT either a mer operon or an integron.";  
RL Mol. Microbiol. 17:1189-1200(1995).

```

RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-identified;
RA Kholodil G.Y.;
RL Russ. J. Genet. 31:1447-1451(1995).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES-Xanthomonas; PLASMID-RPI; TRANSPONSON-TN5053;
RA MEDLINE-93253772; PubMed-8387603;
RL Kholodil G.Y., Yurleva O.V., Lomovskaya O.L., Gorlenko Zh.M.,
  Mindlin S.Z., Nikiforov V.G.;
  "tn5053, a mercury resistance transposon with integron's ends.";
  J. Mol. Biol. 230:1103-1107(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES-Pseudomonas sp. (strain ADP); STRAIN-ADP; PLASMID-PADP-1;
RA Martinez B.M., Tomkins J., Wackett L.P., Wing R., Sadowsky M.T.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC EMBL; X73112; CAAS1538.1; -;
DR EMBL; L20693; AAB05979.1; -;
DR EMBL; Z23095; CAAB0641.1; -;
DR EMBL; L20694; AAB02644.1; -;
DR EMBL; Z23094; CAAB0640.1; -;
DR EMBL; L40585; AAA98322.1; -;
DR EMBL; Z23481; CAAB3889.1; -;
DR EMBL; U66917; AAK50289.1; -;
DR InterPro: IPR000551; HTH_Merr.
DR Pfam: PF00376; merr; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
DR Activator; DNA-binding; Mercuric resistance; Mercury; Plasmid;
  Repressor; Transcription regulation.
SQ SEQUENCE 144 AA; 16060 MW; EAB3FEFEC317F2D CRC64;

Query Match
Best Local Similarity 44.5%; Score 273; DB 2; Length 144;
Matches 53; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 THEEASSLAEHKLDVREKMDLARMETVLSLYCACHARKGNVSCPLIASLGSSGTH 61
DB 80 THEEASSLAEHKLDVREKMDLARMETVLSLYCACHARKGNVSCPLIASLGKKEPR 139
QY 62 CEEA 65
DB 140 SADA 143

RESULT 8
ID 057492 PRELIMINARY; PRT; 151 AA.
AC 057492;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE Regulatory protein.
OS Enterobacter cloacae, and
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
CC NCBI_TaxID=550, 571;
RN [1]
RP SEQUENCE FROM N.A.
RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-SE31;
RC Osborn A.M.;

```

```

RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC EMBL; Z23486; CAAB3894.1; -;
DR EMBL; Z23485; CAAB3893.1; -;
DR InterPro: IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
DR DNA-binding; Transcription regulation.
SQ SEQUENCE 151 AA; 16559 MW; 238460FCE51754AD CRC64;

Query Match
Best Local Similarity 44.0%; Score 269.5; DB 2; Length 151;
Matches 57; Conservative 1; Mismatches 9; Indels 3; Gaps 2;

QY 2 THEEASSLAEHKLDVREKMDLARMETVLSLYCACHARKGNVSCPLIASLGSSGTH 61
DB 80 THEEASSLAEHKLDVREKMDLARMETVLSLYCACHARKGNVSCPLIASLG--DGTK 137
QY 62 CEEASSLAEH 71
DB 138 L-AASARGSH 146

RESULT 9
ID 007304 PRELIMINARY; PRT; 151 AA.
AC 007304;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE MERR protein (Mercuric resistance operon regulatory protein).
GN MERR.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
CC NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TC97;
RX MEDLINE-97303088; PubMed-9159519;
RA Yurleva O., Kholodil G., Minakhin L., Gorlenko Z., Kalyaeva E.,
  Mindlin S., Nikiforov V.;
  "Intercontinental spread of promiscuous mercury-resistance transposons
  in environmental bacteria.";
  Mol. Microbiol. 24:321-329(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TC97;
RA Kholodil G.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kholodil G.Y., Mindlin S.Z., Gorlenko Z.M., Bass I.A., Kalyaeva E.S.,
  Nikiforov V.;
  "Molecular genetic analysis of the tn5041 transposition system.";
  Russ. J. Genet. 36:365-373(2000).
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC EMBL; Y09210; CAAT0409.2; -;
DR EMBL; Y18976; CAB81570.1; -;
DR InterPro: IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
DR DNA-binding; Transcription regulation.
SQ SEQUENCE 151 AA; 16541 MW; 17CC8F1005A33FD0 CRC64;

Query Match
Best Local Similarity 44.0%; Score 269.5; DB 2; Length 151;
Matches 57; Conservative 1; Mismatches 9; Indels 3; Gaps 2;

```

```

QY 2 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQSSGTH 61
DB 80 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQ--DGTK 137
QY 62 CEASSLAEH 71
DB 138 L-AASARGSH 146

RESULT 10
Q57106 PRELIMINARY; PRT; 151 AA.
AC Q57106; 008282; 008130; 008287; 008288; 008185;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Regulatory protein (Mercury resistance operon regulatory protein) (MER operon regulatory protein).
GN MERR.
OS Acinetobacter calcoaceticus,
OS Escherichia coli,
OS Alcaligenes sp.,
OS Pantoea agglomerans,
OS Enterobacter cloacae,
OS Acinetobacter sp., and
OS Acinetobacter sp. LS56-7.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
CC Acinetobacter.
OX NCBI_TaxID=471, 562, 512, 549, 550, 472, 107402;
RN [1]
RP SEQUENCE FROM N.A.
RC Osborn A.M., Bruce K.D., Strick P., Ritchie D.A.;
RL Submitted (Dec-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CH210; TRANSPOSON=TN5059;
RX MEDLINE=97303088; PubMed=9159519;
RA Yurleva O., Kholodil G., Minakhin L., Gorlenko Z., Kalayeva E.,
RA Mindlin S., Nikiforov V.;
RT "Intercontinental spread of promiscuous mercury-resistance transposons in environmental bacteria."
RL Mol. Microbiol. 24:321-329(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Acinetobacter sp.; STRAIN=BW3; PLASMID=PKLH207;
RX TRANSPOSON=TNPKLH207 TNPKLH2-LIKE ABERRANT TRANSPOSON;
RA Kholodil G.Y., Yurleva O.V., Mindlin S.Z., Gorlenko Z.M.,
RA Nikiforov V.G.;
RT "pKH2-like aberrant transposons and possible mechanisms of their dissemination."
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Acinetobacter sp. LS56-7; PLASMID=PKLH204;
RX TRANSPOSON=TNPKLH204;
RA MEDLINE=21272500; PubMed=11376944;
RA Kholodil G.Y.;
RT "The shuffling function of resolvases."
RL Gene 269:121-130(2001).
CC -1 SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL; Y09026; CAAT07040.1; -
DR EMBL; Y08993; CAAT07040.1; -
DR EMBL; Z33483; CAAT07040.1; -
DR EMBL; Z33482; CAAT07040.1; -
DR EMBL; Y08992; CAAT07040.1; -
DR EMBL; Y09025; CAAT07040.1; -
DR EMBL; AJ245842; CAC80722.1; -
DR EMBL; AJ250860; CAC80722.1; -
DR InterPro; IPR000551; HTH_MERR.
DR Pfam; PF00376; merr; 1.

```

```

DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW DNA-binding; Plasmid; transcription regulation.
SQ SEQUENCE 151 AA; 16529 MW; 239350FCE51754AD CRC64;

Query Match
Best Local Similarity 44.0%; Score 269.5; DB 2; Length 151;
Matches 57; Conservative 1; Mismatches 9; Indels 3; Gaps 2;

QY 2 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQSSGTH 61
DB 80 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQ--DGTK 137
QY 62 CEASSLAEH 71
DB 138 L-AASARGSH 146

RESULT 11
Q9R9X0 PRELIMINARY; PRT; 111 AA.
AC Q9R9X0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MERR (Fragment).
GN MERR.
OS Pseudomonas putida.
OC Plasmid group 2 plasmid.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KT2440;
RA Bruce K.D., Lilley A.K., Bailey M.J.;
RT "mer sequences on plasmids."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF092069; AAD52705.1; -
DR InterPro; IPR000551; HTH_MERR.
DR SMART; SM00422; HTH_MERR; 1.
KW Plasmid.
KW NON_TER
FT NON_TER 1 1
FT 111 111
SQ SEQUENCE 111 AA; 12406 MW; 642A1FD89A896C97 CRC64;

Query Match
Best Local Similarity 43.7%; Score 268; DB 2; Length 111;
Matches 52; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQ 55
DB 57 THCEASSLAHEKLDVREKMDLARMETVLSLVACAHARKGNVSCPLIASIQ 110

RESULT 12
Q07300 PRELIMINARY; PRT; 144 AA.
AC Q07300;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Mercuric resistance operon regulatory protein (MERR protein).
GN MERR OR MERR1.
OS Pseudomonas sp., and
OS Pseudomonas sp. ED23-33.
OC Plasmid pMR26.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306, 83781;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Pseudomonas sp.; STRAIN=K-62; PLASMID=pMR26;

```

RX MEDLINE-97311403: PubMed-9168120:  
 RA Kiyono M., Omura T., Inuzuka M., Fujimori H., Pan-Hou H.;  
 RT "Nucleotide sequence and expression of the organomercurial-resistance  
 RT determinants from a Pseudomonas K-62 plasmid pMR26.";  
 RN Gene 189:151-157(1997).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Pseudomonas sp. ED23-33; TRANSPOSON-TN5058;  
 RA Minakhina S., Minakhin L., Kholodil G., Mindlin S., Gorlenko Z.H.,  
 RA Yurleva O., Nikiforov V.;  
 RT Molecular inventory of transposons from environmental bacteria:  
 RT epidemic dissemination of Tn21-, Tn5041-, and Tn5053-related  
 RT transposons.";  
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC EMBL: D83080; BAA20334.1; -;  
 DR EMBL: Y17897; CAC14697.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PRO0040; HTHMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 DR DNA-binding: Plasmid; Transcription regulation.  
 KW SEQUENCE 144 AA; 15651 MW; AE25F7DF73A6734 CRC64;  
 SQ  
 Query Match 43.7%; Score 268; DB 2; Length 144;  
 Best Local Similarity 81.8%; Pred. No. 8.2e-19;  
 Matches 54; Conservative 5; Mismatches 5; Indels 2; Gaps 1;  
 QY 2 THECEASSLAHEKLDVREKMDLARETVLSELYCACHARKGNVSCPLIASLOGSGTH 61  
 DB 80 THECEASSLAHEKLDVREKMDLARETVLSELYCACHARKGNVSCPLIASLOG--GTS 137  
 QY 62 CEEBASS 67  
 DB 138 LAGAST 143  
 RESULT 13  
 P77071 PRELIMINARY; PRT; 144 AA.  
 AC P77071; Q91UN8;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Merr protein (MERR2 protein).  
 GN MERR OR MERR2.  
 OS Escherichia coli.  
 OS Pseudomonas sp. K-62,  
 OS Pseudomonas sp. ED23-33, and  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid R831b, plasmid pMR26, and plasmid pSB102.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 76885, 83781, 382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli;  
 RA MEDLINE-65155497; PubMed-6099319;  
 RA Ogawa H.T., Tolle C.L., Summers A.O.;  
 RT "Physical and genetic map of the organomercury resistance (Omr) and  
 RT inorganic mercury resistance (Hgr) loci of the IncM plasmid R831b.";  
 RL Gene 32:311-320(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli;  
 RA Tolle C., Tolle P., Summers A.O.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli;  
 RA Liebert C.A., Watson A.L., Summers A.O.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Pseudomonas sp. K-62;  
 RA Kiyono M., Hou H.P.;  
 RT "Nucleotide sequence of the two mer operons from a Pseudomonas K-62  
 RT plasmid pMR26.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Pseudomonas sp. ED23-33; TRANSPOSON-TN5058;  
 RA Minakhina S., Minakhin L., Kholodil G., Mindlin S., Gorlenko Z.H.,  
 RA Yurleva O., Nikiforov V.;  
 RT Molecular inventory of transposons from environmental bacteria:  
 RT epidemic dissemination of Tn21-, Tn5041-, and Tn5053-related  
 RT transposons.";  
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-R.meliloti; PLASMID-PSB102;  
 RC TRANSPOSON-MERCURY RESISTANCE TRANSPOSON TN5718;  
 RA Schneider S., Keller M., Droege M., Lanka E., Puehler A.,  
 RA Selitschka W.;  
 RT "The genetic organization and evolution of the broad-host-range  
 RT mercury resistance plasmid pSB102 isolated from a microbial population  
 RT residing in the rhizosphere of alfalfa.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC EMBL: U77087; AAB49638.1; -;  
 DR EMBL: AB013925; BAA36431.1; -;  
 DR EMBL: Y17897; CAC14703.1; -;  
 DR EMBL: AJ304453; CAC79199.1; -;  
 DR InterPro: IPR000551; HTH\_Merr.  
 DR Pfam: PF00376; merr; 1.  
 DR PRINTS: PRO0040; HTHMERR.  
 DR SMART: SM00422; HTH\_MERR; 1.  
 DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
 DR DNA-binding: Plasmid; Transcription regulation.  
 KW SEQUENCE 144 AA; 15623 MW; 9DF4ED08A57D75D CRC64;  
 SQ  
 Query Match 43.7%; Score 268; DB 2; Length 144;  
 Best Local Similarity 81.8%; Pred. No. 8.2e-19;  
 Matches 54; Conservative 5; Mismatches 5; Indels 2; Gaps 1;  
 QY 2 THECEASSLAHEKLDVREKMDLARETVLSELYCACHARKGNVSCPLIASLOGSGTH 61  
 DB 80 THECEASSLAHEKLDVREKMDLARETVLSELYCACHARKGNVSCPLIASLOG--GTS 137  
 QY 62 CEEBASS 67  
 DB 138 LAGAST 143  
 RESULT 14  
 Q44191 PRELIMINARY; PRT; 151 AA.  
 AC Q44191;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Regulatory protein.  
 GN MERR.  
 OS Agrobacterium tumefaciens.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-T217;  
 RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
 RT "Sequence Conservation between Regulatory Mercury Resistance Genes  
 RT from Mercury Polluted and Pristine Environments.";

Submitted (May-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
DR EMBL: 233487; CAB83895.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr; 1.  
DR PRINTS: PR00040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR.1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
DR DNA-binding: Transcription regulation.  
KM SEQUENCE 151 AA; 16591 MW; 548460FCES0240FC CRC64;  
SQ  
Query Match 43.5%; Score 266.5; DB 2; Length 151;  
Best Local Similarity 80.0%; Pred. No. 1.2e-18;  
Matches 56; Conservative 2; Mismatches 9; Indels 3; Gaps 2;  
QY 2 THCEASSLAEHLKDYREKMDLARETYLSELVCACHARKGVSCPLIASLOGSSGTH 61  
DB 80 THCEASGLAEHLKDYREKMDLAREAVLSELVCACHARKGVSCPLIASLO--DGTK 137  
QY 62 CEASSLAEH 71  
DB 138 L-AASARGSH 146  
RESULT 15  
Q99093 PRELIMINARY: PRT; 151 AA.  
AC 099093:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
DE Mer operon regulatory protein.  
GN MERR.  
OS Acinetobacter calcoaceticus.  
OS Acinetobacter lwoffii, and  
OS Acinetobacter sp.  
OG Plasmid pKLH2, Plasmid pKLH103, Plasmid pKLH102, Plasmid pKLH104, and  
OG Plasmid pKLH205.  
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
OC Acinetobacter.  
OX NCBI\_TaxID=471, 28090, 472;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.calcoaceticus; PLASMID=PKLH2;  
RA Lomovskaya O.L., Nikiforov V.G.;  
RL Genetika 24:1064-1071(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.calcoaceticus; PLASMID=PKLH2;  
RA MEDLINE=94134837; PubMed=8302940;  
RA Kholodil G.Y., Lomovskaya O.L., Gorlenko Z.M., Mindlin S.Z.,  
RA Yurleva O.V., Nikiforov V.G.;  
RT "Molecular characterization of an aberrant mercury resistance  
transposable element from an environmental Acinetobacter strain.";  
RL Plasmid 30:303-308(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=A.lwoffii; PLASMID=PKLH103, pKLH102, AND pKLH104;  
RA Kholodil G.Y., Mindlin S.Z., Lomovskaya O.L., Gorlenko Z.M.,  
RA Yurleva O.V., Nikiforov V.G.;  
RT "pKLH1-like aberrant mercury resistance transposons of environmental  
Acinetobacter strains: spread, polymorphism and possible origin.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Acinetobacter sp.; STRAIN=EDA5-25; PLASMID=PKLH205;  
RX MEDLINE=21272500; PubMed=11376944;  
RA Kholodil G.;  
RT "The shuffling function of resolvases.";  
RL Gene 269:121-130(2001).  
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.

DR EMBL: AF213017; AAA19678.1; -;  
DR EMBL: AJ251539; CAB65953.1; -;  
DR EMBL: AJ251517; CAB65939.1; -;  
DR EMBL: AJ250009; CAB65945.1; -;  
DR EMBL: AJ251537; CAB65949.1; -;  
DR EMBL: AJ251706; CAC39408.1; -;  
DR InterPro: IPR000551; HTH\_Merr.  
DR Pfam: PF00376; merr; 1.  
DR PRINTS: PR00040; HTHMERR.  
DR SMART: SM00422; HTH\_MERR.1.  
DR PROSITE: PS00552; HTH\_MERR\_FAMILY; 1.  
DR DNA-binding: Plasmid; Transcription regulation.  
KM SEQUENCE 151 AA; 16561 MW; 549350FCES0240FC CRC64;  
SQ  
Query Match 43.5%; Score 266.5; DB 2; Length 151;  
Best Local Similarity 80.0%; Pred. No. 1.2e-18;  
Matches 56; Conservative 2; Mismatches 9; Indels 3; Gaps 2;  
QY 2 THCEASSLAEHLKDYREKMDLARETYLSELVCACHARKGVSCPLIASLOGSSGTH 61  
DB 80 THCEASGLAEHLKDYREKMDLAREAVLSELVCACHARKGVSCPLIASLO--DGTK 137  
QY 62 CEASSLAEH 71  
DB 138 L-AASARGSH 146  
Search completed: May 28, 2003, 10:28:52  
Job time : 34 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:27:17 : Search time 15 Seconds  
(without alignments)  
229.499 Million cell updates/sec

Title: US-09-977-137A-4  
Perfect score: 613  
Sequence: 1 MTHCEASSLAHKLKDYRE.....HARKGNVSCPSAMSHPOFEK 117

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	15.0	138	4	US-09-134-001C-2982
2	88.5	14.4	299	4	US-09-188-930-332
3	78	12.7	299	4	US-09-188-930-192
4	72	11.7	1388	2	US-08-685-576-4
5	71	11.6	1274	4	US-09-095-443-2
6	70.5	11.5	770	1	US-08-369-796-12
7	70.5	11.5	770	1	US-08-416-581B-1
8	70.5	11.5	770	1	US-08-416-581B-5
9	70.5	11.5	770	1	US-08-416-581B-9
10	70.5	11.5	770	2	US-08-852-091-12
11	70.5	11.5	770	2	US-08-820-754-12
12	70.5	11.5	770	3	US-08-956-652-12
13	70.5	11.5	770	3	US-08-956-869-12
14	70.5	11.5	770	3	US-09-012-710-8
15	70.5	11.5	770	3	US-08-948-547-12
16	70.5	11.5	770	4	US-09-087-465-6
17	70.5	11.5	770	4	US-09-364-970-3
18	70.5	11.5	770	4	US-09-364-970-5
19	70.5	11.5	770	4	US-09-556-273-8
20	70.5	11.5	770	4	US-08-956-653A-12
21	70.5	11.5	770	4	US-09-526-542-2
22	70.5	11.5	770	5	PCT-US95-17025-12
23	70.5	11.5	771	1	US-08-276-099A-14
24	70.5	11.5	771	1	US-08-781-890-14
25	69	11.3	2547	3	US-09-058-489-35
26	68.5	11.2	955	1	US-08-006-676B-1
27	68.5	11.2	955	1	US-08-282-845-2

28	68.5	11.2	955	2	US-08-428-414A-3	Sequence 3, Appl1
29	68.5	11.2	955	5	PCT-US94-00324-1	Sequence 1, Appl1
30	68	11.1	1886	4	US-08-938-105-3	Sequence 3, Appl1
31	66.5	10.8	425	1	US-08-700-749A-5	Sequence 5, Appl1
32	66.5	10.8	425	3	US-09-020-684-5	Sequence 5, Appl1
33	66.5	10.8	425	3	US-09-020-467-5	Sequence 5, Appl1
34	66.5	10.8	425	3	US-09-020-685-5	Sequence 5, Appl1
35	66.5	10.8	425	3	US-09-020-683-5	Sequence 5, Appl1
36	66.5	10.8	574	1	US-08-049-473-2	Sequence 2, Appl1
37	66.5	10.8	574	1	US-08-312-648-2	Sequence 2, Appl1
38	66.5	10.8	574	5	PCT-US94-04190-2	Sequence 2, Appl1
39	66.5	10.8	577	2	US-07-728-215-29	Sequence 29, Appl1
40	66.5	10.8	577	4	US-08-938-085A-29	Sequence 29, Appl1
41	66.5	10.8	783	1	US-08-256-938-2	Sequence 2, Appl1
42	65.5	10.7	386	2	US-08-559-303B-75	Sequence 75, Appl1
43	65.5	10.7	386	4	US-09-175-828-75	Sequence 75, Appl1
44	65.5	10.7	659	4	US-08-781-891-77	Sequence 77, Appl1
45	65.5	10.7	810	2	US-08-820-170A-34	Sequence 34, Appl1

ALIGNMENTS

```
RESULT 1
US-09-134-001C-2982
; Sequence 2982, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2982
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2982

Query Match      15.0%; Score 92; DB 4; Length 138;
Best Local Similarity 21.8%; Pred. NO. 0.0036;
Matches 27; Conservative 16; Mismatches 37; Indels 44; Gaps 2;

QY 28 METVLSLVCAC-----HARKGNVSCP----- 49
||:|||||
DB 4 MEKKISELAKACVKNKETVYERKGLIACPPNRESGRYISEPTADRVRFIRKMKELDF 63
||:|||||

QY 50 -----LIASLOGSSSTHCEASSLAHKLKDYREKMAIARMETVLSLVCACHARKN 103
||:|||||
DB 64 SLKEHLLRGVDDDERCKDMYAFVQKTKFERKVDLRIORLLELKEKCPDERKAI 123
||:|||||

QY 104 VSCP 107
||
DB 124 YTCP 127

RESULT 2
US-09-188-930-332
; Sequence 332, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
```

;; TITLE OF INVENTION: and Methods For Their Use  
;; FILE REFERENCE: 11000.1011cl  
;; CURRENT APPLICATION NUMBER: US/09/188,930A  
;; CURRENT FILING DATE: 1998-11-09  
;; NUMBER OF SEQ ID NOS: 348  
;; SOFTWARE: FASTSEQ for Windows Version 3.0  
;; SEQ ID NO 332  
;; LENGTH: 299  
;; TYPE: PRF  
;; ORGANISM: Mouse  
US-09-188-930-332

Query Match 14.4%; Score 88.5; DB 4; Length 299;  
Best Local Similarity 27.5%; Pred. No. 0.025;  
Matches 25; Conservative 2; Mismatches 25; Indels 39; Gaps 4;

QY 38 ACHARKNVSCPLASLOGSGTHCEEA-----SLAEHLKDVREKMDLARMETVLS 91  
DB 41 ACDPVSGTCICP-----FGKTGCHGRCGCPDRFGKGCHEK----- 76  
QY 92 ELVCA-----CHARKGVSCPSAMSHPOFE 116  
DB 77 ---CACRNGGLCHATNGSCSCPLGMWMPHCE 104

RESULT 3  
US-09-188-930-192  
; Sequence 192; Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE REFERENCE: 11000.1011cl  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 192  
; LENGTH: 299  
; TYPE: PRF  
; ORGANISM: mouse  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (98)...(98)  
; NAME/KEY: UNSURE  
; LOCATION: (239)...(239)  
US-09-188-930-192

Query Match 12.7%; Score 78; DB 4; Length 299;  
Best Local Similarity 24.1%; Pred. No. 0.38;  
Matches 28; Conservative 5; Mismatches 25; Indels 58; Gaps 6;  
QY 3 HCEEA-----SLAEHLKDVREKMDLARMETVLS 50  
DB 59 HCEGCGCPDRFGKGCHEK-----CACRNGGLCHATNGSCSCPL 96  
QY 51 IASLOGSGTHCEEAASSLAEHLKDVREKMDLARMETVLS 106  
DB 97 -----GXMGPCHCEHACPAGRYGACATLE-----CSC-----QNNGSC 128

RESULT 4  
US-08-685-576-4  
; Sequence 4; Application US/08685576  
; Patent No. 5906819  
; GENERAL INFORMATION:  
; APPLICANT: Kaiduchi, Kozo  
; APPLICANT: Iwamatsu, Akihito

;; APPLICANT: Nakano, Takeshi  
;; APPLICANT: Ito, Masaaki  
;; APPLICANT: Takahashi, No. 5906819uaki  
;; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
;; NUMBER OF SEQUENCES: 16  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 3000 K Street, N.W., Suite 500  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20007-5109  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/685,576  
;; FILING DATE: 24-JUL-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 7-325129  
;; FILING DATE: 20-NOV-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 8-17150  
;; FILING DATE: 05-JAN-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 8-131206  
;; FILING DATE: 26-APR-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bent, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 16887/843  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1388 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-685-576-4

Query Match 11.7%; Score 72; DB 2; Length 1388;  
Best Local Similarity 25.0%; Pred. No. 13;  
Matches 30; Conservative 22; Mismatches 28; Indels 40; Gaps 5;

QY 4 CEEASSLA-----EHKLDVREKMA-----DLAR-----METVLS 33  
DB 887 CEEKTLGKELQOKKQELDERDSLANQLETTTKADSEQLARSLAEQYSDLEKEKIMK 946  
QY 34 EE-----VCACHARKGVSCPLASLOGSGTHCEEAASLA-----EHKLDVREKMDL 83  
DB 947 EEIKEMARHKQELTEKATIASLEETINRTLTSDVANLANEKEELNKLADVQPOL 1006

RESULT 5  
US-09-095-443-2  
; Sequence 2; Application US/09095443  
; Patent No. 6342593  
; GENERAL INFORMATION:  
; APPLICANT: Plozman, Gregory  
; APPLICANT: Peles, Etor  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
; TITLE OF INVENTION: OF ALP RELATED DISORDERS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700



GENERAL INFORMATION:  
 APPLICANT: Kishimoto, Tadamiitsu  
 APPLICANT: Akira, Shizuo  
 TITLE OF INVENTION: TRANSCRIPTION FACTOR APPE  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SUGIYUE, MION, ZINN, MACPEAK & SEAS  
 STREET: 2100 Pennsylvania Avenue, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/416,581B  
 FILING DATE: 04-APR-1995  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: JP 6-565825/1994  
 FILING DATE: 04-APR-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Nakamura, Dean H.  
 REGISTRATION NUMBER: 33,981  
 REFERENCE/DOCKET NUMBER: O-37891  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)293-7060  
 TELEFAX: (202)293-7860  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 770 amino acids



CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,091  
FILING DATE: 06-MAY-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/369,796  
FILING DATE: 06-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-852-091-12

Query Match 11.5%; Score 70.5; DB 2; Length 770;  
Best Local Similarity 23.3%; Pred. No. 9.2; Mismatches 38; Indels 21; Gaps 4;  
Matches 24; Conservative 20;

QY 5 EASLSAEHKLKDV-----EKADLARMETVLSVLCACHARKGN-VSCPLIASIQ 55  
DB 73 QESNVLVYHNRIRKQFLOSRYLEKPMETAR-----IVARCLWEESRLDQTAATAAQ 125  
QY 56 GSSGTH-----CEASLSAEHKLKDVREKADLARMETVLSL 93  
DB 126 GGOANHPTAAVTEKQMLEQHLQDVRRKRVODLEQMKRYVENL 168

RESULT 11  
US-08-820-754-12  
Sequence 12, Application US/08820754  
Patent No. 5976835  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
NUMBER OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/820,754  
FILING DATE: 19-MAR-1997

CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-820-754-12

Query Match 11.5%; Score 70.5; DB 2; Length 770;  
Best Local Similarity 23.3%; Pred. No. 9.2; Mismatches 38; Indels 21; Gaps 4;  
Matches 24; Conservative 20;

QY 5 EASLSAEHKLKDV-----EKADLARMETVLSVLCACHARKGN-VSCPLIASIQ 55  
DB 73 QESNVLVYHNRIRKQFLOSRYLEKPMETAR-----IVARCLWEESRLDQTAATAAQ 125  
QY 56 GSSGTH-----CEASLSAEHKLKDVREKADLARMETVLSL 93  
DB 126 GGOANHPTAAVTEKQMLEQHLQDVRRKRVODLEQMKRYVENL 168

RESULT 12  
US-08-956-652-12  
Sequence 12, Application US/08956652  
Patent No. 6013475  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
NUMBER OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,652  
FILING DATE: 19-MAR-1997  
CLASSIFICATION:

PRIOR APPLICATION DATA: 08/212,185  
APPLICATION NUMBER: US 08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-652-12

Query Match 11.5%; Score 70.5; DB 3; Length 770;  
Best Local Similarity 23.3%; Pred. No. 9.2;  
Matches 24; Conservative 20; Mismatches 38; Indels 21; Gaps 4;  
QY 5 EASSLAEHLKLDVR-----EKADLARMTVLSLYCACHARKGN-VSCPLIASIQ 55  
DB 73 QESNVLYQHNLRRKQFLOSRYLEKPEIAR-----IVARCIWESRLLOTATATAAQ 125  
QY 56 GSSGTH-----CEASSLAEHLKLDVREKADLARMTVLSL 93  
DB 126 GGOANHPTAAVTEKQMLEQHLQDVAKRVODLEQKMKVENVL 168

RESULT 13  
US-08-956-869-12  
Sequence 12, Application US/08956869  
Patent No. 6030808  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,869  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/212,185  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-869-12

Query Match 11.5%; Score 70.5; DB 3; Length 770;  
Best Local Similarity 23.3%; Pred. No. 9.2;  
Matches 24; Conservative 20; Mismatches 38; Indels 21; Gaps 4;  
QY 5 EASSLAEHLKLDVR-----EKADLARMTVLSLYCACHARKGN-VSCPLIASIQ 55  
DB 73 QESNVLYQHNLRRKQFLOSRYLEKPEIAR-----IVARCIWESRLLOTATATAAQ 125  
QY 56 GSSGTH-----CEASSLAEHLKLDVREKADLARMTVLSL 93  
DB 126 GGOANHPTAAVTEKQMLEQHLQDVAKRVODLEQKMKVENVL 168

RESULT 14  
US-09-012-710-8  
Sequence 8, Application US/09012710  
Patent No. 6087478  
GENERAL INFORMATION:  
APPLICANT: Vinkemeier, Uwe  
APPLICANT: Marelli, Ismail  
APPLICANT: Kuriyan, John  
APPLICANT: Darnell, Jr., James E.  
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A  
TITLE OF INVENTION: STAT PROTEIN AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,710  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-194  
TELECOMMUNICATION INFORMATION:





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2003, 10:28:57 ; Search time 56 Seconds  
(without alignments)  
207.200 Million cell updates/sec

Title: US-09-977-137A-4  
Sequence: 1 MTHCEASLAHKLKDVRE.....HARKGNVSCPSAMSHPOFEK 117

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :  
1: /pubseq/ptoddata/1/pubseq/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptoddata/1/pubseq/PTCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptoddata/1/pubseq/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptoddata/1/pubseq/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptoddata/1/pubseq/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptoddata/1/pubseq/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptoddata/1/pubseq/PTCT\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptoddata/1/pubseq/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptoddata/1/pubseq/US08\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptoddata/1/pubseq/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptoddata/1/pubseq/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptoddata/1/pubseq/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptoddata/1/pubseq/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptoddata/1/pubseq/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.5	14.4	296	US-09-866-050A-458	Sequence 458, App
2	88.5	14.4	299	US-09-866-050A-192	Sequence 192, App
3	88.5	14.4	299	US-09-866-050A-332	Sequence 332, App
4	71.5	11.7	337	US-10-259-042-22	Sequence 22, Appl
5	71	11.6	245	US-09-864-761-34088	Sequence 34088, A
6	71	11.6	474	US-10-084-994-11	Sequence 11, Appl
7	71	11.6	868	US-09-884-001-19	Sequence 19, Appl
8	71	11.6	1274	US-10-020-215-2	Sequence 2, Appl
9	71	11.6	1725	US-10-037-182-12	Sequence 12, Appl
10	71	11.6	1786	US-10-037-182-10	Sequence 10, Appl
11	71	11.6	1786	US-09-938-275-7	Sequence 7, Appl
12	71	11.6	2383	US-10-082-830-260	Sequence 260, App
13	70.5	11.5	497	US-09-796-753-116	Sequence 116, App
14	70.5	11.5	770	US-10-045-792-8	Sequence 8, Appl
15	70.5	11.5	770	US-10-038-010-56	Sequence 56, Appl
16	70.5	11.5	793	US-09-925-302-780	Sequence 780, App
17	70	11.4	689	US-10-108-605-305	Sequence 305, App
18	69.5	11.3	241	US-09-987-107-24	Sequence 24, Appl
19	69	11.3	310	US-10-084-994-12	Sequence 12, Appl

20	69	11.3	337	US-09-800-362-2	Sequence 2, Appl
21	69	11.3	366	US-09-938-803-13	Sequence 13, Appl
22	69	11.3	1507	US-10-043-487-330	Sequence 330, App
23	66.5	10.8	577	US-10-072-841-29	Sequence 29, Appl
24	66	10.8	116	US-09-864-761-44770	Sequence 44770, A
25	66	10.8	934	US-09-842-758-18	Sequence 18, Appl
26	65.5	10.7	386	US-09-753-143-75	Sequence 75, Appl
27	65.5	10.7	810	US-09-976-165-34	Sequence 34, Appl
28	65.5	10.7	1480	US-10-011-064-5	Sequence 5, Appl
29	64.5	10.5	732	US-09-381-624A-3	Sequence 3, Appl
30	64	10.4	428	US-09-738-626-5780	Sequence 5780, Ap
31	64	10.4	561	US-09-922-364A-1	Sequence 1, Appl
32	64	10.4	561	US-09-254-590-1	Sequence 1, Appl
33	64	10.4	561	US-10-115-695-1	Sequence 1, Appl
34	64	10.4	561	US-10-115-651-1	Sequence 1, Appl
35	64	10.4	561	US-10-115-671-1	Sequence 1, Appl
36	64	10.4	561	US-10-115-415-1	Sequence 1, Appl
37	64	10.4	561	US-10-115-260-1	Sequence 1, Appl
38	63.5	10.4	111	US-10-091-504-917	Sequence 917, App
39	63.5	10.4	111	US-09-764-869-917	Sequence 917, App
40	63.5	10.4	194	US-09-866-050A-335	Sequence 335, App
41	63.5	10.4	586	US-10-092-390-4	Sequence 4, Appl
42	63.5	10.4	636	US-09-796-753-100	Sequence 100, App
43	63.5	10.4	636	US-09-796-753-124	Sequence 124, App
44	63.5	10.4	796	US-10-028-392-5	Sequence 5, Appl
45	63.5	10.4	1140	US-10-092-390-2	Sequence 2, Appl

## ALIGNMENTS

```
RESULT 1
US-09-866-050A-458
; Sequence 458, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Oornst, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 458
; LENGTH: 296
; TYPE: PRT
; ORGANSIM: Mouse
US-09-866-050A-458

Query Match      14.4%; Score 88.5; DB 9; Length 296;
Best Local Similarity 27.5%; Pred. No. 0.23;
Matches 25; Conservative 2; Mismatches 25; Indels 39; Gaps 4;

OY 38 ACIARKGNVSCPLIASIGSSGTHCEEA-----SLAEHLKDVREKMDIAREMYLVS 91
DB 38 ACDPVSGTCTCP-----PGKTGGHCEGRCPPDRFGKCEHR----- 73
OY 92 ELVCA-----CHARKGNVSCPSAMSHPOFE 116
DB 74 ----CACRNGSLCHATNIGSCSCPLGMMGPHE 101

RESULT 2
US-09-866-050A-192
; Sequence 192, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-192

Query Match          14.4%; Score 88.5; DB 9; Length 299;
Best Local Similarity 27.5%; Pred. No. 0.23;
Matches 25; Conservative 2; Mismatches 25; Indels 39; Gaps 4;

QY 38 ACHARKNVSCPLIASLOGSSGTHCEA-----SLAEHLKIDVREKMDLARMETVLS 91
Db 41 ACDPVSGTICP-----PGKTGHCERCPCPDREKGCERK----- 76

QY 92 ELVCA-----CHARKNVSCPSAMSHPOFE 116
Db 77 ---CACRNGGLCHATNGSCSCPJGWMGPHE 104

RESULT 3
US-09-866-050A-332
; Sequence 332, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-332

Query Match          14.4%; Score 88.5; DB 9; Length 299;
Best Local Similarity 27.5%; Pred. No. 0.23;
Matches 25; Conservative 2; Mismatches 25; Indels 39; Gaps 4;

QY 38 ACHARKNVSCPLIASLOGSSGTHCEA-----SLAEHLKIDVREKMDLARMETVLS 91
Db 41 ACDPVSGTICP-----PGKTGHCERCPCPDREKGCERK----- 76

QY 92 ELVCA-----CHARKNVSCPSAMSHPOFE 116
Db 77 ---CACRNGGLCHATNGSCSCPJGWMGPHE 104

RESULT 4
US-10-259-042-22
; Sequence 22, Application US/10259042
; Publication No. US20030077644A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Yang, Bing
; TITLE OF INVENTION: Diagnosis and Treatment of Diseases Caused by Mutations in CD7
; FILE REFERENCE: SAG 0105 PUS
; CURRENT APPLICATION NUMBER: US/10/259,042
; CURRENT FILING DATE: 2002-09-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-259-042-22

Query Match          11.7%; Score 71.5; DB 9; Length 337;
Best Local Similarity 24.0%; Pred. No. 14;
Matches 25; Conservative 16; Mismatches 46; Indels 17; Gaps 3;

QY 16 KDVERKMDLARMETVSELVACACHARKNVSCPLIASLOGSSGTH-----CEAS 66
Db 117 QOLREKISQLOKREVELOK-----ARKELISSQ--DTIQEKQTHEDAEQOLQACQAE 168

QY 67 SLAEHLKIDVREKMDLARMETVSELVACACHARKNVSCPSAW 110
Db 169 AKTKENLKTEERERRDLDRLTSTRETLRRFSDSSDTCPCGW 212

RESULT 5
US-09-864-761-34088
; Sequence 34088, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
```



PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34088  
LENGTH: 245  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000215.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.9  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 11  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.8  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11  
OTHER INFORMATION: EST HUMAN HT: BF375387.1, EVALUATE 2.00e-37  
OTHER INFORMATION: SWISSPROT HT: O95613, EVALUATE 1.00e-107  
US-09-864-761-34088

Query Match 11.6%; Score 71; DB 10; Length 245;  
Best Local Similarity 23.3%; Pred. No. 11;  
Matches 31; Conservative 18; Mismatches 38; Indels 46; Gaps 8;

4 CEESSLAIEH---KLDVREKMDL-ARMETVSELVCACHARKGNVSCPLIASLOGSS 58  
91 CVMODYAHALLQKLEKRSYVDLQMLEKVOQ---ALHSQO-----QLEADA 138  
59 GTHCEASSLAIEHKLDVREKMDLARMETVSELVCACHARKGNVSC-----PSAM- 110  
139 QKCEALR-----RER-----EVSATLKSTVEALHTQKRELKSLEREREKPAWL 183  
111 -----SHPOFEK 117  
184 QAELEQSHPRLEK 196

RESULT 6  
US-10-084-994-11  
Sequence 11, Application US/10084994  
Publication No. US20030023070A1  
GENERAL INFORMATION:  
APPLICANT: NI et al.  
TITLE OF INVENTION: Attractin-like Polynucleotides, Polypeptides, and Antibodies  
FILE REFERENCE: P011P1C1  
CURRENT APPLICATION NUMBER: US/10/084,994  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/790,621  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: PCT/US00/23663  
PRIOR FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: US 60/151,348  
PRIOR FILING DATE: 1999-08-30  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn Ver. 3.1  
SEQ ID NO 11  
LENGTH: 474  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-084-994-11

Query Match 11.6%; Score 71; DB 9; Length 474;  
Best Local Similarity 26.7%; Pred. No. 23;  
Matches 24; Conservative 8; Mismatches 24; Indels 34; Gaps 5;

38 ACHARKGNVSCPLIASLOGSGTHCEEA---SSLAEHKLKDVREKMDLARMETVSELV 94

DB 115 ACHVYTGACRCP-----DFTGSGCEQACPPSGFGE-----DCAQM----- 150  
QY 95 C-----ACHARKGNVSCPSAMSHPOFEK 117  
151 CQCPENPACHPATGTCCSCACGHPSCQO 180

RESULT 7  
US-09-884-001-19  
Sequence 19, Application US/09884001  
Publication No. US20020182656A1  
GENERAL INFORMATION:  
APPLICANT: Bird, Timothy A.  
APPLICANT: Peschon, Jacques J.  
APPLICANT: Sims, John E.  
APPLICANT: Virca, G. Duke  
APPLICANT: Willis, Cynthia R.  
TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF  
FILE REFERENCE: Immunex GNR/SGNK PCT  
CURRENT APPLICATION NUMBER: US/09/884,001  
CURRENT FILING DATE: 2001-06-18  
PRIOR APPLICATION NUMBER: 60/113,003  
PRIOR FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 868  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-884-001-19

Query Match 11.6%; Score 71; DB 9; Length 868;  
Best Local Similarity 27.1%; Pred. No. 48;  
Matches 32; Conservative 17; Mismatches 57; Indels 12; Gaps 4;

5 EEASSLAIEHKLDVREKMDLARMETVSELVCACHARKGNVSCPL-----IASLQ-GS 57  
437 KELSADQMEILLROEYKEKADFLAEOQLLELASHSHTEQQLRASLMAQEAQLOJRL 496  
58 SGTCEASSLAIEHKLDVREKMDLARMETVSELVCACHARK-----GNVSCPSAM 110  
497 RSTESQLEALAEQPPQNOAQOALASIVSALQOALGVSYESRPELSGGDSAPSV 554

RESULT 8  
US-10-020-215-2  
Sequence 2, Application US/10020215  
Publication No. US20030008347A1  
GENERAL INFORMATION:  
APPLICANT: PLOWMAN, GREGORY  
APPLICANT: PELES, EIOR  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF ALP RELATED DISORDERS  
FILE REFERENCE: 038602/1290  
CURRENT APPLICATION NUMBER: US/10/020,215  
CURRENT FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: 09/095,443  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/049,477  
PRIOR FILING DATE: 1997-06-11  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1274  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: ALP  
OTHER INFORMATION: polypeptide sequence  
US-10-020-215-2

Query Match 11.6%; Score 71; DB 9; Length 1274;  
Best Local Similarity 24.6%; Pred. No. 77;



TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-082-830-260

Query Match  
Best Local Similarity 27.1%; Pred. No. 16e+02;  
Matches 32; Conservative 17; Mismatches 57; Indels 12; Gaps 4;

11.6%; Score 71; DB 9; Length 2383;

QY 5 EBASLAHKKLVREKADLARMETVLSLVCAHAKGNSCP-----IASIQ-GS 57  
DB 1072 KELSQMELLROVEKEKADFLAEOQLLEELASHITTEQLRLASIMAOEAKAQLQLRL 1131

QY 58 SGTHEEASSLAHKKLVREKADLARMETVLSLVCAHAKGNSCP-----GNVSCPSAW 110  
DB 1132 RSTESQLEALAEQDPGNQAOAQLASLSALQALGSCVCSRPPLSGGDSAPSVW 1189

RESULT 13  
US-09-796-753-116  
Sequence 116, Application US/09796753  
Publication No. US2003002798A1  
GENERAL INFORMATION:  
APPLICANT: McCarthy, Sean A.  
TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
FILE REFERENCE: 7853-227-999  
CURRENT APPLICATION NUMBER: US/09/796,753  
PRIORITY FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 09/183,175  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 09/223,094  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/223,546  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/224,246  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/259,388  
PRIOR FILING DATE: 1999-02-26  
PRIOR APPLICATION NUMBER: 60/122,458  
PRIOR FILING DATE: 1999-03-01  
PRIOR APPLICATION NUMBER: 09/312,359  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/336,536  
PRIOR FILING DATE: 1999-06-18  
PRIOR APPLICATION NUMBER: 09/342,687  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: 09/345,464  
PRIOR FILING DATE: 1999-06-30  
PRIOR APPLICATION NUMBER: 09/365,164  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/399,723  
PRIOR FILING DATE: 1999-09-20  
PRIOR APPLICATION NUMBER: 09/409,634  
PRIOR FILING DATE: 1999-09-30  
PRIOR APPLICATION NUMBER: 09/471,179  
PRIOR FILING DATE: 1999-12-23  
PRIOR APPLICATION NUMBER: 09/474,071  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 09/474,072  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 09/514,010  
PRIOR FILING DATE: 2000-02-25  
PRIOR APPLICATION NUMBER: 09/516,745  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/572,002  
PRIOR FILING DATE: 2000-05-14  
PRIOR APPLICATION NUMBER: 09/597,993  
PRIOR FILING DATE: 2000-06-19  
PRIOR APPLICATION NUMBER: 09/599,596  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/630,334  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: 09/606,565  
PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: 09/606,317  
PRIOR FILING DATE: 2000-06-29  
PRIOR APPLICATION NUMBER: 09/665,666  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: 09/677,751  
PRIOR FILING DATE: 2000-09-30  
NUMBER OF SEQ ID NOS: 162  
SEQ ID NO 116  
LENGTH: 497  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-796-753-116

Query Match  
Best Local Similarity 11.5%; Score 70.5; DB 9; Length 497;  
Matches 22; Conservative 6; Mismatches 28; Indels 25; Gaps 4;

QY 38 ACHARKGNVSCPLIASLGSSGTHCEASSLAHKKLVREKADLARMETVLSLV-C- 95  
DB 91 SCHPSDGTGSC-----LAGWTGPDCSEACPGHMGK-----CSOLQGH 130

QY 96 ---ACHARKGNVSCPSAWSH 113  
DB 131 HGCTCHPDGSCICTPGWTGP 151

RESULT 14  
US-10-045-792-8  
Sequence 8, Application US/10045792  
Publication No. US20030003563A1  
GENERAL INFORMATION:  
APPLICANT: Vinkemeier, Dwe  
Darnell, Jr., James E.  
Kutliyan, John  
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A  
STAT PROTEIN AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Krauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/045,792  
FILING DATE: 19-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,710  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-194  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO

SEQUENCE	DESCRIPTION:	SEQ ID NO:	8:
US-10-045-792-8			

Query Match	11.5%	Score 70.5;	DB 9;	Length 770;
Similarity	23.3%	Pred. 4.7;		
Best Local	24;	Mismatches	38;	Indels 21;
Matches	24;	Conservative		Gaps 4.

```

09 5 EASSLAEHKLKDV-----EKMDLARMETVLSLVACCHARKN-VSCPLIASLQ 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 QESNVLQHNLRRIQFLQSRYLEKPMELAR-----IVARCLMEESRLQTAATAAQ 125

```

```

Qy 56 GSSGTH-----CEASSLAERKLVREKMAIDARMTVLSL 93
      | : | : | : | : | : | : | : | : | : | : | : |
Db 126 GGQANHPTAAVYTEKQMLEQHLQDVRRKRVQDLEQKKMKEVNI 168

```

RESULT 15  
RE-10-038

Sequence 56, Application US/10038010  
Publication No. US20030040089A1

GENERAL INFORMATION:

APPLICANT: Pierre, Legrain

**TITLE OF INVENTION:** Protein-protein interactions in adipocyte cells

FILE REFERENCE: B4767A

CURRENT FILING DATE: 2002-07-23

PRIOR APPLICATION NUMBER: US 60/259,377

PRIOR FILING DATE: 2001-01-02

NUMBER OF SEQ ID NOS: 67

SOFTWARE: PatentIn version 3.1  
 GEO ID NO 56

```
SEQ ID NO 56
LENGTH: 770
```

LENGTH: 770  
TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: STAT3 : Transcription factor

LOCATION: (1)..(770)  
OTHER INFORMATION:

OTHER INFORMATION:  
TS-10-038-010-56

Query Match	11.58; Score 70.5; DB 9; Length 770;
-------------	--------------------------------------

Best Local Similarity	Conservative	Mismatches	Indels	Gaps
24	20	38	21	4

SECRET

5 EEA5SLAEHKLKDV-----EKMADLARMETVLSLVCACHARKGN-VSCPLIASLQ 55

```

      :|:: | :| :: || ::|| :| : : | :|
77 OECUUT VOUHNT PBTROET OSBPVI EKPDMETAB-----TYVBCI WEESRLJOTATAA00 1235

```

73 QESVLYQHNLRRIKQFLOSRYLEKPMELAR-----IVARCLWE

56 GSSGTH-----CEEASSLAHHKLDVREKMDIARMTVLSSEL 93

36 GSSGIB-----CEHSSJAHNUNDNVKNENTUJRWVIAVJOCZ 35

126 GGQANHPTAAVLTKEQOMLEQHLQDVKKVQDLEQKKVENVL 168

Search completed: May 28, 2003, 10:38:14

Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 05:03:23 ; Search time 1798 seconds

(without alignments)  
5195.773 Million cell updates/sec

Title: US-09-977-137A-3

Perfect score: 321

Sequence: 1 atgacacactgcgagagagc.....agggaatgttctctgcccg 321

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_ov: \*  
7: gb\_ov: \*  
8: gb\_ov: \*  
9: gb\_ov: \*  
10: gb\_ov: \*  
11: gb\_ov: \*  
12: gb\_ov: \*  
13: gb\_ov: \*  
14: gb\_ov: \*  
15: gb\_ov: \*  
16: gb\_ov: \*  
17: gb\_ov: \*  
18: gb\_ov: \*  
19: gb\_ov: \*  
20: gb\_ov: \*  
21: gb\_ov: \*  
22: gb\_ov: \*  
23: gb\_ov: \*  
24: gb\_ov: \*  
25: gb\_ov: \*  
26: gb\_ov: \*  
27: gb\_ov: \*  
28: gb\_ov: \*  
29: gb\_ov: \*  
30: gb\_ov: \*  
31: gb\_ov: \*  
32: gb\_ov: \*  
33: gb\_ov: \*  
34: gb\_ov: \*  
35: gb\_ov: \*  
36: gb\_ov: \*  
37: gb\_ov: \*  
38: gb\_ov: \*  
39: gb\_ov: \*  
40: gb\_ov: \*  
41: gb\_ov: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	100.0	321	6	AX473152
2	321	100.0	354	6	AX473166
3	321	100.0	509	6	AX473167
4	164.4	51.2	435	6	AX473150
5	164.4	51.2	445	1	PFL422224
6	164.4	51.2	445	1	PFL422225
7	164.4	51.2	3763	1	NRI1004
8	164.4	51.2	5747	1	R1004
9	164.4	51.2	19672	1	AF071413
10	164.4	51.2	26652	1	AP188331
11	164.4	51.2	94281	1	AP000342
12	164.4	51.2	218160	1	STYPPHOM1
13	164.4	50.7	516	1	ARMERGE4
14	145.2	45.2	516	1	ARMERGE8
15	140.6	43.8	333	1	AF092069
16	137.2	42.7	516	1	PEMERGE10
17	137.2	42.7	516	1	PEMERGE9
18	137.2	42.7	1340	1	TRN501
19	137.2	42.7	6769	1	PS080214
20	137.2	42.7	8355	1	ISTN501
21	137.2	42.7	221851	1	AF348706
22	136.6	42.6	331	1	AF092070
23	136.6	42.6	537	1	PEMERGE1
24	136.6	42.6	1012	15	ASTNMERR
25	136.6	42.6	1012	1	PEMERGE30A
26	136.6	42.6	1928	1	PFTNMERR
27	136.6	42.6	3355	1	PLMER05A
28	136.6	42.6	108845	1	PEMER
29	135.8	42.3	537	1	ARMERGE5
30	135.8	42.3	3977	1	PSU90263
31	135	42.1	9566	1	RPIW17A
32	135	42.1	116580	1	PU344068
33	134.2	41.8	537	1	ACMERGE2
34	134.2	41.8	1040	1	ECMERGE6
35	134.2	41.8	1040	1	ECMERGE1
36	134.2	41.8	2695	1	PSMERTP
37	134.2	41.8	3150	1	PSPT18976
38	134.2	41.8	3356	1	ASP250860
39	134.2	41.8	4256	1	ASP245842
40	134.2	41.8	4824	1	AF461013
41	134.2	41.8	8012	1	ECMERTREL
42	132.6	41.3	537	1	ACMERGE3
43	132.6	41.3	537	1	ARMERGE7
44	132.6	41.3	537	1	ARMERGE7
45	132.6	41.3	537	1	ARMERGE7

#### ALIGNMENTS

RESULT 1  
AX473152  
LOCUS AX473152  
DEFINITION Sequence 3 from Patent WO0230962.  
ACCESSION AX473152  
VERSION AX473152.1 GI:22207871  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 Summers, A.O. and Cagliati, J.J.  
Metal binding proteins, recombinant host cells and methods  
Patent: WO 0230962-A 3 18-APR-2002;  
UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)

Pred. No. is the number of results predicted by chance to have a

FEATURES  
source  
Location/Qualifiers  
1..321  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chelon"  
BASE COUNT 72 a 94 c 103 g 52 t  
ORIGIN  
Query Match 100.0%; Score 321; DB 6; Length 321;  
Best Local Similarity 100.0%; Pred. No. 2.3e-60;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACACACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAG 60  
DB 1 ATGACACACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAG 60  
QY 61 AAGATGGCGCACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
DB 61 AAGATGGCGCACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
QY 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
DB 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
QY 181 CACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAGATG 240  
DB 181 CACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAGATG 240  
QY 241 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 300  
DB 241 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 300  
QY 301 AAGGGGAATGTTTCTGCGCCG 321  
DB 301 AAGGGGAATGTTTCTGCGCCG 321  
RESULT 2  
AX473166 354 bp DNA linear PAT 09-AUG-2002  
LOCUS AX473166  
DEFINITION Sequence 17 from Patent W00230962.  
ACCESSION AX473166  
VERSION AX473166.1 GI:22207876  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 Summers,A.O. and Caguiat,J.J.  
AUTHORS Metal binding proteins, recombinant host cells and methods  
TITLE Patent: WO 0230962-A 17 18-APR-2002;  
JOURNAL UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)  
FEATURES  
source  
Location/Qualifiers  
1..354  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="sequence encoding chelon"  
BASE COUNT 83 a 103 c 111 g 57 t  
ORIGIN  
Query Match 100.0%; Score 321; DB 6; Length 354;  
Best Local Similarity 100.0%; Pred. No. 2.3e-60;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACACACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAG 60  
DB 1 ATGACACACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAG 60  
QY 61 AAGATGGCGCACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
DB 61 AAGATGGCGCACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
QY 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
DB 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120

DB 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 180  
QY 181 CACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAGATG 240  
DB 181 CACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAGATG 240  
QY 241 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 300  
DB 241 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 300  
QY 301 AAGGGGAATGTTTCTGCGCCG 321  
DB 301 AAGGGGAATGTTTCTGCGCCG 321  
RESULT 3  
AX473167 509 bp DNA linear PAT 09-AUG-2002  
LOCUS AX473167  
DEFINITION Sequence 18 from Patent W00230962.  
ACCESSION AX473167  
VERSION AX473167.1 GI:22207877  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 Summers,A.O. and Caguiat,J.J.  
AUTHORS Metal binding proteins, recombinant host cells and methods  
TITLE Patent: WO 0230962-A 18 18-APR-2002;  
JOURNAL UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)  
FEATURES  
source  
Location/Qualifiers  
1..509  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="sequence encoding chelon flanked by sequences  
derived from plasmid"  
BASE COUNT 139 a 128 c 140 g 102 t  
ORIGIN  
Query Match 100.0%; Score 321; DB 6; Length 509;  
Best Local Similarity 100.0%; Pred. No. 2.2e-60;  
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACACACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAG 60  
DB 138 ATGACACACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAG 197  
QY 61 AAGATGGCGCACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 120  
DB 198 AAGATGGCGCACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCAT 257  
QY 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 180  
DB 121 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 180  
QY 258 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 317  
DB 258 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 317  
QY 318 CACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAGATG 240  
DB 318 CACTGGAGAGAGGCGCAGCAGCTGCGGCAACAAGCTCAAGAGCTGCGGAGATG 240  
QY 241 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 300  
DB 378 GCGGACTTGGCGGCGCATGAAACCGTGTCTGAACCTGTGTGCGCTGCCATGACGA 437  
QY 301 AAGGGGAATGTTTCTGCGCCG 321  
DB 438 AAGGGGAATGTTTCTGCGCCG 458  
RESULT 4  
AX473150 435 bp DNA linear PAT 09-AUG-2002  
LOCUS AX473150  
DEFINITION Sequence 1 from Patent W00230962.

ACCESSION AX473150  
VERSION AX473150.1 GI:22207870  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1  
AUTHORS Summers, A.O. and Caguiat, J.J.  
TITLE Metal binding proteins, recombinant host cells and methods  
JOURNAL Patent: WO 0230967-A 1 18-APR-2002;  
UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)  
FEATURES  
source  
1. .435  
/organism="Shigella flexneri"  
/db\_xref="taxon:623"  
/note="Tn21 of Plasmid R100"  
BASE COUNT 95 a 113 c 146 g 81 t  
ORIGIN

Query Match 51.2%; Score 164.4; DB 6; Length 435;  
Best Local Similarity 96.6%; Pred. No. 3.3e-26;  
Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 ACACACTGCGAGAGAGCCAGCAGCTGGCCGAAACACAGCTCAAGAGCGTGGCGAGAG 63  
DB 238 ACCCATGTCGAGGAGGAGCCAGCAGCTGGCCGAAACACAGCTCAAGAGCGTGGCGAGAG 297  
OY 64 ATGGCCGACTTGGCGCGATGGAACCGTGTCTGAGACTGTGCGCCCTCCATGCA 123  
DB 298 ATGGCCGACTTGGCGCGATGGAACCGTGTCTGAGACTGTGCGCCCTCCATGCA 357  
OY 124 CGAAGGGGAATGTTCTCTGCGCCGTTGATCGGCTACTACAGGATCTCAGGC 177  
DB 358 CGAAGGGGAATGTTCTCTGCGCCGTTGATCGGCTACTACAGGAGGCGAAGCAGGC 411

RESULT 5  
PFL422224/c 445 bp DNA linear BCT 07-JUL-2002  
LOCUS PFL422224  
DEFINITION Pseudomonas fluorescens (strain KHP22) transposon Tn5041g, partial  
sequence (region of insertion of Tn21 homologue).  
ACCESSION AJ422224  
VERSION AJ422224.1 GI:19848211  
KEYWORDS meir gene; Meir protein.  
SOURCE Pseudomonas fluorescens.  
ORGANISM Pseudomonas fluorescens  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.

REFERENCE  
1  
AUTHORS Kholodil, G.Y., Gorlenko, Z.M., Mindlin, S.Z. and Nikiforov, V.G.  
TITLE Distribution of distinct microvariants of Tn5041 in environmental  
bacteria  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 445)  
AUTHORS Kholodil, G.Y.  
TITLE Direct Submission  
JOURNAL Submitted (18-DEC-2001) Kholodil G.Y., Russian Academy of Sciences,  
Institute of Molecular Genetics, Kurchatov Sq. 2, Moscow 123182,  
RUSSIA  
COMMENT  
Related sequence Y09209.  
FEATURES  
source  
1. .445  
Location/Qualifiers  
/organism="Pseudomonas fluorescens"  
/strain="KHP22"  
/db\_xref="taxon:294"  
/sex="Tn21 (narrow range)"  
/Plasmid="PKHP22"  
/country="Kyrgyzstan:Khaidarkan"  
/note="Isolated from mercury mine, ore"  
repeat\_region 1. .445  
repeat\_unit 31. .35

/note="the sequence that has been duplicated in the course  
of insertion of the Tn21 homologue"  
/rpt\_type=DIRECT  
36. .445  
repeat\_region /note="Tn21 homologue taking the integron"  
36. .73  
repeat\_unit /transposon="Tn21deltaTn2"  
/note="Tn21 subgroup transposons"  
/rpt\_type=INVERTED  
/rpt\_type=TERMINAL  
complement(69. .445)  
/gene="meir"  
/pseudo  
complement(69. .445)  
/gene="meir"  
/function="activator/repressor"  
/pseudo  
/codon\_start=3  
/transl\_table=11  
/product="Meir protein"  
BASE COUNT 88 a 149 c 119 g 89 t  
ORIGIN

Query Match 51.2%; Score 164.4; DB 1; Length 445;  
Best Local Similarity 96.6%; Pred. No. 3.3e-26;  
Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 ACACACTGCGAGAGAGCCAGCAGCTGGCCGAAACACAGCTCAAGAGCGTGGCGAGAG 63  
DB 266 ACCCATGTCGAGGAGGAGCCAGCAGCTGGCCGAAACACAGCTCAAGAGCGTGGCGAGAG 207  
OY 64 ATGGCCGACTTGGCGCGATGGAACCGTGTCTGAGACTGTGCGCCCTCCATGCA 123  
DB 206 ATGGCCGACTTGGCGCGATGGAACCGTGTCTGAGACTGTGCGCCCTCCATGCA 147  
OY 124 CGAAGGGGAATGTTCTCTGCGCCGTTGATCGGCTACTACAGGATCTCAGGC 177  
DB 146 CGAAGGGGAATGTTCTCTGCGCCGTTGATCGGCTACTACAGGAGGCGAAGCAGGC 93

RESULT 6  
PFL422225/c 445 bp DNA linear BCT 07-JUL-2002  
LOCUS PFL422225  
DEFINITION Pseudomonas fluorescens (strain KHP25) transposon Tn5041g, partial  
sequence (region containing a Tn21 homologue).  
ACCESSION AJ422225  
VERSION AJ422225.1 GI:19848212  
KEYWORDS meir gene; Meir protein.  
SOURCE Pseudomonas fluorescens.  
ORGANISM Pseudomonas fluorescens  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.

REFERENCE  
1  
AUTHORS Kholodil, G.Y., Gorlenko, Z.M., Mindlin, S.Z. and Nikiforov, V.G.  
TITLE Distribution of distinct microvariants of Tn5041 in environmental  
bacteria  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 445)  
AUTHORS Kholodil, G.Y.  
TITLE Direct Submission  
JOURNAL Submitted (18-DEC-2001) Kholodil G.Y., Russian Academy of Sciences,  
Institute of Molecular Genetics, Kurchatov Sq. 2, Moscow 123182,  
RUSSIA  
COMMENT  
Related sequence Y09209.  
FEATURES  
source  
1. .445  
Location/Qualifiers  
/organism="Pseudomonas fluorescens"  
/strain="KHP25"  
/db\_xref="taxon:294"  
/sex="Tn21 (narrow range)"  
/Plasmid="PKHP22"  
/country="Kyrgyzstan:Khaidarkan"  
/note="Isolated from mercury mine, ore"  
repeat\_region 1. .445  
repeat\_unit 31. .35

```

repeat_unit
/feature="Tn5041G"
31..35
/note="the sequence that has been duplicated in the course
of insertion of the Tn21 homologue"
/rpe_type=DIRECT
36..>445
/note="Tn21 homologue lacking the integron"
/transposon="Tn21deltaIn2"
36..73
/note="Tn21 subgroup transposons"
/rpe_type=INVERTED
/rpe_type=TERMINAL
complement(69..445)
/gene="merR"
complement(69..>445)
/gene="merR"
/function="activator/repressor"
/codon_start=3
/transl_table=11
/product="MerR protein"
/protein_id="CAD19592.1"
/db_xref="GI:19848213"
/translation="VETIRYORKGLPEPDKPYGSIRRYGEADVVRVYKSAQRLG
ESIDETAEILRLDDGTHCEASSLAEHKLVDREKMDIAMETVLSLVYVCAHARKG
NVSCPLIASLOGEAGLARSAMP"

BASE COUNT      88 a      149 c      119 g      89 t
ORIGIN
Query Match      51.2%; Score 164.4; DB 1; Length 445;
Best Local Similarity 96.6%; Pred. No. 3.3e-26;
Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

OY 4 ACACACTGCGAGAGAGCCACACAGCCCTGCCGACAACAACGACAGCGCGGAGAG 63
    |||||||
DB 266 ACCCACTGGAGAGAGCGACACAGCTGCGACAACAAGCTGACAGCTGCGCGAGAG 207
    |||||||
OY 64 ATGGCCGACTTGCGCGCATGAAACCGTGTCTGTGACTGCTGTGCGCTGCCATGCA 123
    |||||||
DB 206 ATGGCCGACTTGCGCGCATGAAACCGTGTCTGTGACTGCTGTGCGCTGCCATGCA 147
    |||||||
OY 124 CGAAGAGGAGATGTTCTGCTGATGCGCGTGTGATGCGCGTGTGCGCGTGTGCG 177
    |||||||
DB 146 CGAAGAGGAGATGTTCTGCTGATGCGCGTGTGATGCGCGTGTGCGCGTGTGCG 93
    |||||||

RESULT 7
NR1MER/c      3763 bp      DNA      linear      BCT 16-FEB-1994
LOCUS      NR1mercury resistance (mer) operon.
DEFINITION      KO3089
ACCESSION      KO3089.1 GI:150389
VERSION      merA gene; merR gene; mercuric reductase.
KEYWORDS      Plasmid NR1 (IncFII) from E.coli DNA, clone pDB7.
SOURCE      Plasmid NR1
ORGANISM      Plasmids.
1 (bases 1 to 3763)
Barrineau,P., Gilbert,P., Jackson,W.J., Jones,C.S., Summers,A.O.
and Wisdom,S.
The DNA sequence of the mercury resistance operon of the IncFII
Plasmid NR1
J. Mol. Appl. Genet. 2 (6), 601-619 (1984)

JOURNAL      MEDLINE
PUBMED      85159407
6530603
2 (bases 1 to 3763)
Summers,A.O.
Unpublished (1986)
[2] revises [1].
Draft entry and sequence in computer readable form for [2] kindly
provided by A.O. Summers, 15-SEP-1986. Potential Shine-Dalgarno
sequences are located at positions 157-162 (13.1 kd), 590-596
(12.4 kd), 1740-1745 (59 kd) and 545-542 (15.9 kd). A single 38 bp
inverted repeat, which delimits the leftward end of Tn21, Tn4 and
Tn501 extends from nucleotide 65-102. A promoter region for the

```

structural genes is located between positions 366-585. This area contains regions of dyad symmetry which might function as operator sites for merR, which negatively regulates the expression of the structural genes as well as its own expression. The proteins deduced from the open reading frames all have counterparts in other organisms. The 12.4 kd and 15.0 kd predicted peptides are located in the inner membrane and have no detectable precursors. The 9.4 kd peptide is soluble, is processed to a 12 kd form and is for the most part associated with the periplasm. The first 30 bases were replaced with the corrected sequence from [2]. The rest of the revisions are in the sites table.

```

FEATURES
source
CDS
1..3763
/organism="Plasmid NR1"
/db_xref="taxon:2467"
/plasmid="Plasmid NR1"
complement(98..532)
/note="15.9 kd merR protein"
/codon_start=1
/transl_table=11
/protein_id="AAB59072.1"
/db_xref="GI:455296"
/translation="MENNLENTLITGVFAKAGVNETIRYORKGLRPPDKPYGSIR
RGADVYRKVFYSACRIGSDIEAELRLDDGTHCEASSLAEHKLVDREKMD
IARMETVLSLVYVCAHARKGVSCPLIASLOGEAGLARSAMP"
179..541
/note="13.1 kd protein"
/codon_start=1
/transl_table=11
/protein_id="AAB59073.1"
/db_xref="GI:150391"
/translation="MAGAHROHGFHAROVGHLLAHVLEIVFGAAGLLAVAIYEP
QQLNLVOTEPQPLCFRHFHFPHNKLPLAADAAVYLVFPOQALLIENDCLHYDPC
RLGRNANGVFOIIFHIA"
216..598
/note="6.5 kd protein"
/codon_start=1
/transl_table=11
/protein_id="AAB59074.1"
/db_xref="GI:150392"
/translation="WRAKSAIRSRSTSLCSARILLASSOWPSSSRNSAISRLNS
RCADFTNPTRTTSASP"
604..954
/note="12.4 kd merTA protein"
/codon_start=1
/transl_table=11
/protein_id="AAB59075.1"
/db_xref="GI:150393"
/translation="MSEPONGRGALEFAGLIAIILASTCCGIPVLVALFGSGANIL
TVLEPYRPLEIGALVALFPAWKRIRYPOACKPGEVCAIPQVATYKLLFWIVAVLV
LVALGFPPVPPVFFY"
968..1243
/note="9.4 kd merTB protein"
/codon_start=1
/transl_table=11
/protein_id="AAB59076.1"
/db_xref="GI:150394"
/translation="MKKLFASIALAAVAPVMAATQVTLAVPGMTCAACPIYVKAL
SKVEGTVKYDVGFEKREAVYTFPDITASVQKILKATPDADAPSSVKQ"
1279..1701
/note="15.0 kd merTC protein"
/codon_start=1
/transl_table=11
/protein_id="AAB59077.1"
/db_xref="GI:150395"
/translation="KGLMTRIAKTKGALGSVVSAMGCAACAPALASGALIGIFLSQ
YEGFLISRLPLFALALFLANALGWSHROMLSLGMIGAIYFAATVWLLGNMWTIA
NLMYVGIALMIGVISINDFVSPAHRRCCPDCELEPAKRL"
1753..3447
/note="Hg(II) reductase (59.0 kd merA)"
/codon_start=1
/transl_table=11
/protein_id="AAB59078.1"

```





gene 1080..1262  
/gene="mer"  
CDS 1080..1262  
/gene="mer"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA092260.1"  
/db\_xref="GI:151749"  
/translation="MRKSAIFSRTSLISLCARILLASSOWPVSSSRNSAISRLNPS  
RCADPTNFRRTWSASP"  
1468..1818  
/gene="mer"  
/gene="mer"  
1468..1818  
/gene="mer"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA092261.1"  
/db\_xref="GI:151750"  
/translation="MSEPPNGRGALEFAGGLAAILASTCCIGLPIVVALGFGAWTGNL  
TVLEPRPLFTGALVALFPAWKRTIPVQCKGVECAIPQVRAITYLITPIYAVLV  
LVALGPPYVPPFY"  
1832..2107  
/gene="mer"  
/gene="mer"  
1832..2107  
/gene="mer"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA092262.1"  
/db\_xref="GI:151751"  
/translation="MKLFASLALAAVAPYMAATQVTVLAVPGMTCAAPTYYKAL  
SVEGSKYDVGVGEKREAVVTFDDTKASVCKLTATADAGPSSVKQ"  
2182..2184  
/citation={3}  
2617..4311  
/gene="mer"  
2617..4311  
/gene="mer"  
/codon\_start=1  
/transl\_table=11  
/product="mercuric reductase"  
/protein\_id="AA092263.1"  
/db\_xref="GI:151752"  
/translation="MSTLKITGMCDCSCAVHVKALERKVGVSADVSYAKSALAI  
EVGSGPALAAVAGATLADAPSVTPGGLDKMDLGRNKTGSSGALHAY  
IGSGGMAAALAKAVEGARGVTLERGTGGCVNMGVPSKIMIAAHABRRSP  
FDGIAATPTTIOATLALAOOARVDELPHAKVEGILEENPAITVYHSGARFENRL  
IVOLNDGGERVAFDRCLITAGASRAVPPIBCLKPTPYTSEALVSEITPKRLAYIG  
SSVVALDLAFAFARIGAVTILASTLFREDDPAIGEAYTAFAFMGCIERYETQASQ  
VAVYINBGDGEFVLTAGELRADKLVAATGAPNTRKIALDNTGVTLLTVOGALVDP  
GMRTSVSHIYAAGDCTDQPOFVYVAAAGTAAINMTGGDALNLTAMPVAVTFDPV  
ATVGYSEAEAHHDGKIDSRLLTLDNVRALNFDTRGFIKLVVEGSGSLIGQAAV  
PAGELIOTATLAIARNMTVOELADQLPPLIMVEGLKLAAGTFNNDVQLSCCAG"  
4329..4691  
/gene="mer"  
/gene="mer"  
4329..4691  
/gene="mer"  
/note="putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA092264.1"  
/db\_xref="GI:151753"  
/translation="MSAYTSQLAHNAAGVSHIVRDYLVKGLRPVACTGGCGVEDD  
AALQICFVRAAFEGIGLIDALFARLCRALDADAGAAQALAVVOLLVERRRALAIL  
DAOLASMAERHHEALP"  
BASE COUNT 1105 a 1775 c 1771 g 1096 t  
ORIGIN  
Query Match 51.2%; Score 164.4; DB 1; Length 5747;  
Best Local Similarity 96.6%; Pred. No. 2.9e-26;  
Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 4 ACACACTCGAGAGCCAGCCTGGCCGACACAGCAAGCTCAAGACGTCGGCAGAG 63

DB 1159 ACCACTGCAGAGAGGCGACAGCCTGGCCGACACAGCAAGCTCAAGAGAGTGGCAGAG 1100  
QY 64 ATGGCCGACTTGGCGCCGATGAAACCGTCTGTCTGTAACCTGTGGCCATGCA 123  
DB 1099 ATGGCCGACTTGGCGCCGATGAAACCGTCTGTCTGTAACCTGTGGCCATGCA 1040  
QY 124 CGAAGGGGAATGTTCTGCGCCGTTGATCGCTACACAGGATCTCAGGC 177  
DB 1039 CGAAGGGGAATGTTCTGCGCCGTTGATCGCTACACAGGAGGAGCAGGC 986  
RESULT 9  
AF071413  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AF071413  
19672 bp DNA linear BCF 13-AUG-2002  
Assembled sequence of transposon Tn21 containing transposition  
genes, the *Integron* *Int2*, and the mercury resistance operon.  
AF071413  
AF071413.2 GI:21930298  
Escherichia coli.  
Escherichia coli.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia  
1 (bases 2952 to 4066)  
Diver,W.P., Grinstead,J., Fritzing,D.C., Brown,N.L.,  
Altenbuchner,J., Rogowsky,P. and Schmidt,R.  
DNA sequences of and complementation by the *tupr* genes of Tn21,  
Tn501 and Tn1721  
Mol. Gen. Genet. 191 (2), 189-193 (1983)  
84013495  
6312271  
2 (bases 18411 to 19672)  
Mistra,T.K., Brown,N.L., Fritzing,D.C., Pridmore,R.D.,  
Barnes,W.M., Haberstroh,L. and Silver,S.  
Mercuric ion-resistance operons of plasmid R100 and transposon  
Tn501: the beginning of the operon including the regulatory region  
and the first two structural genes  
Proc. Natl. Acad. Sci. U.S.A. 81 (19), 5975-5979 (1984)  
85014891  
6091128  
3 (bases 15973 to 19672)  
Barrineau,P., Gilbert,P., Jackson,W.J., Jones,C.S., Summers,A.O.  
and Wisdom,S.  
The DNA sequence of the mercury resistance operon of the IncFII  
plasmid NRI  
J. Mol. Appl. Genet. 2 (6), 601-619 (1984)  
85159407  
6530603  
4 (bases 16253 to 18549)  
Mistra,T.K., Brown,N.L., Haberstroh,L., Schmidt,A., Goddette,D. and  
Silver,S.  
Mercuric reductase structural genes from plasmid R100 and  
transposon Tn501: functional domains of the enzyme  
Gene 34 (2-3), 253-262 (1985)  
85232071  
2989109  
5 (bases 2952 to 4066)  
Rogowsky,P., Hallford,S.E. and Schmidt,R.  
Definition of three resolvase binding sites at the *res* loci of Tn21  
and Tn1721  
EMBO J. 4 (8), 2135-2141 (1985)  
86055774  
2998784  
6 (bases 3057 to 4105)  
Hyde,D.R. and Tu,C.P.  
tupr: a novel regulatory gene that enhances Tn21 transposition and  
suppresses colinear integrate resolution  
Cell 42 (2), 629-638 (1985)  
85282620  
2992807  
7 (bases 14935 to 15434)  
Brown,N.L., Mistra,T.K., Winnie,J.N., Schmidt,A., Self,M. and

TITLE	Silver, S.	AUTHORS	Recchia, G.D., Stokes, H.W. and Hall, R.M.
JOURNAL	The nucleotide sequence of the mercuric resistance operons of	TITLE	Characterisation of specific and secondary recombination sites
MEDLINE	Plasmid R100 and transposon Tn501: further evidence for mer genes	JOURNAL	Recognised by the integron DNA integrase
PUBMED	which enhance the activity of the mercuric ion detoxification	MEDLINE	Nucleic Acids Res. 22 (11), 2071-2078 (1994)
REFERENCE	system	PUBMED	94301790
AUTHORS	Mol. Gen. Genet. 202 (1), 143-151 (1986)	REFERENCE	80290014
JOURNAL	8 (bases 1 to 3056)	AUTHORS	18 (bases 4048 to 4159; 14945 to 15031)
MEDLINE	Ward, E. and Grinstead, J.	TITLE	Radstrom, P., Skold, O., Swedberg, G., Flensburg, J., Roy, P.H. and
PUBMED	The nucleotide sequence of the tnpA gene of Tn21	JOURNAL	Sundstrom, L.
REFERENCE	Nucleic Acids Res. 15 (4), 1799-1806 (1987)	MEDLINE	Transposon Tn5090 of plasmid R751, which carries an integron, is
AUTHORS	87146495	PUBMED	related to Tn7, Mu, and the retroelements
TITLE	3029727	JOURNAL	J. Bacteriol. 176 (11), 3257-3268 (1994)
JOURNAL	9 (bases 4106 to 7537)	MEDLINE	8195081
MEDLINE	Sundstrom, L., Radstrom, P., Swedberg, G. and Skold, O.	REFERENCE	19 (bases 7540 to 9981)
PUBMED	Site-specific recombination promotes linkage between trimethoprim-	AUTHORS	Hall, R.M., Brown, H.J., Brookes, D.E. and Stokes, H.W.
REFERENCE	and sulfonamide resistance genes. Sequence characterization of	TITLE	Integrons found in different locations have identical 5' ends but
AUTHORS	dhfrv and sulI and a recombination active locus of Tn21	JOURNAL	variable 3' ends
MEDLINE	Mol. Gen. Genet. 213 (2-3), 191-201 (1988)	MEDLINE	J. Bacteriol. 176 (20), 6286-6294 (1994)
PUBMED	89039710	PUBMED	95014074
JOURNAL	3054482	REFERENCE	7929000
MEDLINE	10 (bases 5384 to 5397; 6199 to 6253)	AUTHORS	20 (bases 4492 to 8992)
REFERENCE	Schmidt, F.R., Nucken, E.J. and Henschke, R.B.	TITLE	Recchia, G.D. and Hall, R.M.
AUTHORS	Structure and function of hot spots providing signals for	JOURNAL	Gene cassettes: a new class of mobile element
MEDLINE	site-directed specific recombination and gene expression in Tn21	MEDLINE	Microbiology 141 (Pt 12), 3015-3027 (1995)
PUBMED	transposons	PUBMED	96118227
JOURNAL	Mol. Microbiol. 3 (11), 1545-1555 (1989)	REFERENCE	8574395
MEDLINE	90136084	AUTHORS	21 (bases 9951 to 14934)
PUBMED	2559298	TITLE	Brown, H.J., Stokes, H.W. and Hall, R.M.
JOURNAL	11 (bases 7540 to 9257)	MEDLINE	The integrons In0, In2, and In5 are defective transposon
REFERENCE	Stokes, H.W. and Hall, R.M.	AUTHORS	derivatives
MEDLINE	A novel family of potentially mobile DNA elements encoding	JOURNAL	J. Bacteriol. 178 (15), 4429-4437 (1996)
PUBMED	site-specific gene-integration functions: Integrons	MEDLINE	96345606
REFERENCE	Mol. Microbiol. 3 (12), 1669-1683 (1989)	REFERENCE	8755869
JOURNAL	2560119	AUTHORS	22 (bases 5343 to 6256)
MEDLINE	12 (bases 6301 to 6333)	TITLE	Collis, C.M., Kim, M.J., Stokes, H.W. and Hall, R.M.
PUBMED	Guertineau, F., Brooks, L. and Mullineaux, P.	JOURNAL	Binding of the purified integron DNA integrase Int11 to integron-
REFERENCE	Expression of the sulfonamide resistance gene from plasmid R46	MEDLINE	and cassette-associated recombination sites
JOURNAL	Plasmid 23 (1), 35-41 (1990)	PUBMED	Mol. Microbiol. 29 (2), 477-490 (1998)
MEDLINE	90272799	REFERENCE	9720866
PUBMED	2190244	AUTHORS	23 (bases 5343 to 5404)
JOURNAL	13 (bases 4106 to 5438)	TITLE	Gravel, A., Fournier, B. and Roy, P.H.
REFERENCE	Merger, J., Lachapelle, J., Couture, F., Lafond, M., Vezina, G.,	AUTHORS	DNA complexes obtained with the integron integrase Int11 at the
MEDLINE	Boisinstout, M. and Levesque, R.C.	JOURNAL	attII site
PUBMED	Structural and functional characterization of tnpI, a recombinase	MEDLINE	Nucleic Acids Res. 26 (19), 4347-4355 (1998)
REFERENCE	J. Bacteriol. 172 (7), 3745-3757 (1990)	PUBMED	98416190
JOURNAL	2163386	REFERENCE	9742234
MEDLINE	14 (bases 14967 to 15674)	AUTHORS	24 (bases 1 to 19672)
PUBMED	Allmeier, H., Crennar, B., Greck, M. and Schmitt, R.	TITLE	Liebert, C.A., Hall, R.M. and Summers, A.O.
REFERENCE	Complete nucleotide sequence of Tn1721: gene organization and a	JOURNAL	Transposon Tn21, flagship of the floating genome
JOURNAL	novel gene product with features of a chemotaxis protein	MEDLINE	Microbiol. Mol. Biol. Rev. 63 (3), 507-522 (1999)
MEDLINE	Gene 111 (1), 11-20 (1992)	PUBMED	99407129
PUBMED	92192465	AUTHORS	10477306
REFERENCE	1312499	REFERENCE	25 (bases 7540 to 14934)
JOURNAL	15 (bases 19146 to 19199)	AUTHORS	Partridge, S.R. and Hall, R.M.
MEDLINE	Summers, A.O.	TITLE	In34, a complex In5-like class 1 integron containing orf513 and
PUBMED	Unwist and shout: a heavy metal-responsive transcriptional	JOURNAL	dfnA10
REFERENCE	regulator	MEDLINE	Unpublished
JOURNAL	J. Bacteriol. 174 (10), 3097-3101 (1992)	REFERENCE	correction of sequence errors
MEDLINE	92250400	AUTHORS	26 (bases 4106 to 7539)
PUBMED	1577681	TITLE	Sundstrom, L.
REFERENCE	16 (bases 5134 to 5281)	JOURNAL	Direct Submission
AUTHORS	Levesque, C., Brassard, S., Lapointe, J. and Roy, P.H.	REMARK	Submitted (27-NOV-1989)
TITLE	Diversity and relative strength of tandem promoters for the	JOURNAL	Genbank Accession Number X12870
JOURNAL	antibiotic-resistance genes of several integrons	REFERENCE	27 (bases 7540 to 14934)
MEDLINE	Gene 142 (1), 49-54 (1994)	AUTHORS	Hall, R.M.
PUBMED	94237488	TITLE	Direct Submission
REFERENCE	8181756	Query Match	51.2%; Score 164.4; DB 1; Length 19672;
JOURNAL	17 (bases 4492 to 8992)	Best Local Similarity	96.6%; Pred. No. 2.7e-26;
PUBMED		Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	

[illegible]

gene  
LNTVTAIYALMNTVYLERKAHAALRGNAHVADUSLDDIYSELDGEMHINLDDYIDNRKMGV  
KTAGCFRPLRLOPA"  
complement(3003. .3563)  
/gene="tmpr"  
complement(3003. .3563)  
/gene="tmpr"  
/codon\_start=1  
/transl\_table=11  
/product="putative resolvase"  
/protein\_id="AAG14403.1"  
/db\_xref="GI:10185691"  
/translation="MTGGRIGYIRVSEFDONPEROLEGVKVDFAESDKAGDKVRPO  
LEALISFARCGD/VVHSMRRLANLDDLRIVOTLQREGVHLEFVKEHLSFGEOSP  
MANIMSWKAFEFERALTREORRESITALAKRGAYRGKKSLSSRIALAEHQVEA  
GEOKTLAREFGISRETLVQLRDTQ"  
complement(4242. .5255)  
/gene="inti"  
complement(4242. .5255)  
/gene="inti"  
/codon\_start=1  
/transl\_table=11  
/product="putative integrase"  
/protein\_id="AAG14404.1"  
/db\_xref="GI:10185692"  
/translation="MKTATAPLPILRSVKVLDQLERIRIYLSLPTEQAYVHWVRAP  
IRFHVGHNPATIGSEVFAEFLSWLANERKYSYHROALALAEFYKVLCTDPLMQLQ  
ETGRRPSRPLPVYLTPDEYVRIILGFEGEHRLEPAOLLYGMRISGDLRYVEDDPL  
DHGTTIYREGKSGKDRALMLPESTAPLRDLSRANMWMLKDOAEAGVGLPALPRLER  
KYPAGHSWPEWFVFAQHTHSTDPRSGVRRHHNYDOTFOAEFRRVAGVGITPAPR  
HTLHRSHATALLRSGYDIRTYVDLLGHSVSTWITVTHLVGAGVGRSPDLALPILT  
SER"  
7600. .8439  
/gene="sulr"  
7600. .8439  
/gene="sulr"  
/function="may confer resistance to sulphonamide"  
/codon\_start=1  
/transl\_table=11  
/product="dihydropteroate synthase"  
/protein\_id="AAG14405.1"  
/db\_xref="GI:10185693"  
/translation="MTVYEGTILNTEDSFDESRRLDPGAVTAAIEMLRVSGDVVDV  
GPAAHPHARVSPADEIRLRIAPLLDASDOMHRSIDSPQROKRYLAKRGVGLND  
IQGFPEALPYDIAEADCRILAVHMSAQDGLATIRGHLRRPDALDEIVREPEERSAL  
VRSQAAARLLILDPGMGFMSPAETSLHVSLNOKLKSAGLPLIVSVSKSPUGAG  
VGLPKYDGPASIALAEHLAIGNADYVRTAHPGDLRSATFSETLAKFRSRDARGG  
DHA"  
complement(14272. .15177)  
/gene="mphA"  
complement(14272. .15177)  
/gene="mphA"  
/function="may confer resistance to erythromycin"  
/codon\_start=1  
/transl\_table=11  
/product="macrolide 2'-phosphotransferase"  
/protein\_id="AAG14406.1"  
/db\_xref="GI:10185694"  
/translation="MTVYTTADTSQLYALAAARGKIAGPLTVNELGSDYRVIAIVT  
DQRNWLRIIPRAEVSAAVEPARVYLAIAKNNLPRAVPDPMVGAANEIVAYPMLDESTF  
AVIQGSSPPWVPODSEVFAESFATLAALHAVPISADVAGMLIRTPQAROMV  
AVTVRRARFEVNVDRILRHMQRMLLDDSSMPFSVYHGDIVGVLLDNTERVSGM  
DMSGARVDPAIDMAAHLNVFGKEGIALKLLITYTAGGRVMPRLAHHAIERLAFGAVV  
VALPALDSENEYVLAARKQOLAAE"  
16153. .16968  
/gene="aphI"  
16153. .16968  
/gene="aphI"  
/gene="aphI"  
/function="confers resistance to kanamycin"  
/codon\_start=1  
/transl\_table=11  
/product="aminoglycoside 3'-phosphotransferase"  
/protein\_id="AAG14407.1"  
/db\_xref="GI:10185695"





## JOURNAL

## COMMENT

Submitted (25-Oct-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:  
Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.

## FEATURES

## source

(URL, <http://www.sanger.ac.uk/Projects/S-typhi/>).

Location/Qualifiers

1. 218160

/organism="Salmonella enterica subsp. enterica serovar Typhi"

/strain="CT18"

/db\_xref="taxon:90370"

/plasmid="pHCM1"

/gene="HCM1.01c"

/complement(1..528)

/partial

/gene="HCM1.01c"

/note="HCM1.01c, possible membrane protein, len: 185 aa; unknown function, contains hydrophobic, possible membrane-spanning regions. Spans sequence end"

/codon\_start=1

/transl\_table=11

/product="putative membrane protein"

/protein\_id="CAD09619.1"

/db\_xref="GI:16505741"

/translation="MNQETGSGFTETTVDEERLIRYMGIDVGRYKRPDPDRQ KATFVNGGLMALILALMLICIIORLATGRIMSGFGEVDYRDIOTFLMIVTERL RLVYFLIATGEFSRPMLYRLIPAKPVMFLFDEVEDADVLSSNPFIKRLHKEITIKI RLIVYRLDEERLIRYMGIDVGRYKRPDPDRQ"

/complement(536..540)

/note="possible RBS"

/complement(742..1053)

/gene="HCM1.02c"

/complement(742..1053)

/gene="HCM1.02c"

/note="HCM1.02c, hypothetical protein, len: 103 aa; unknown function"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAD09620.1"

/db\_xref="GI:16505742"

/translation="MAYPDITGVKKKLSOLEVGATWMSLSRTKMGNTTIKIVTLHSV IKEYHDHNVIASWNGNAPRRFGTAITGKKREKPLIRDRSGSARLATREERARILDT K"

/complement(1367..1690)

/gene="HCM1.03c"

/complement(1367..1690)

/gene="HCM1.03c"

/note="HCM1.03c, hypothetical protein, len: 107 aa; unknown function"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAD09621.1"

/db\_xref="GI:16505743"

/translation="MKNPLVSFAELSSNAINAVASVIDMEMDATREKIGKARSLEHS GHRAYNGIFLIQSAANQIAVIRKLGDTRKLDACITENLCMFSEGYLYLFMQRRI NEVVA"

/complement(1697..1702)

/note="possible RBS"

/complement(1871..2585)

/note="ISI"

1871..1883

/note="13 bp inverted repeat flanking ISI"

/complement(1876..2346)

/gene="insB"

/note="HCM1.04c"

/complement(1876..2346)

/gene="insB"

/note="HCM1.04c, insB, possible ISI transposase, len: 156

aa: highly similar to many from Enterobacteriaceae e.g. SW:ISB\_ECOLI (EMBL:D10483), insB, Escherichia coli insertion element ISI protein insB (167 aa), fasta scores: E(): 0, 91.6% identity in 167 aa overlap"

/codon\_start=1

/transl\_table=11

/product="putative ISI transposase"

/protein\_id="CAD09622.1"

/db\_xref="GI:16505744"

/translation="MPCNRPYGRWPHDTSFKRLRPOSVTSRIQPSDYICAKSR ORMLFYADRLKRTVVAHVGEPTATLRISLSPDYVMTGMPYERLKGK LHVISKRYTORERHNLNROHLARIGKRSLSFSKVEJLHDKVIGHYLNKHYQ"

/complement(2265..2540)

/gene="insA"

/note="insA"

/complement(2265..2540)

/note="HCM1.05c, insA, probable ISI transposase, len: 91 aa; highly similar to many from Enterobacteriaceae e.g. SW:ISAL\_ECOLI (EMBL:X52534), insA, Escherichia coli insertion element ISI protein insA (91 aa), fasta scores: E(): 0, 98.9% identity in 91 aa overlap. Contains P500017 ATP/GTP-binding site motif A (P-loop). Contains probable helix-turn-helix motif at aa 67-88"

/codon\_start=1

/transl\_table=11

/product="putative ISI transposase"

/protein\_id="CAD09623.1"

/db\_xref="GI:16505745"

/translation="WASVISCPSGCSATGCVYRNGKSTAGHORVLCSHOKRTWQFT YVASOPGTHOKIIMANGVGRATARIMGVGLNLRILRLKSGSR"

/complement(2284..2289)

/gene="insA"

/note="possible translational frameshift site, similar to that determined experimentally (EMBL:X52534)"

/complement(2472..2495)

/gene="insA"

/note="P500017 ATP/GTP-binding site motif A (P-loop)"

/complement(2573..2585)

/note="13 bp inverted repeat flanking ISI"

/complement(2694..3644)

/gene="corA"

/note="HCM1.07c"

/complement(2694..3644)

/gene="corA"

/note="HCM1.07c, corA, probable magnesium and cobalt transport protein, len: 316 aa; similar to many e.g. SW:COR\_A\_ECOLI (EMBL:L11042), corA, Escherichia coli magnesium and cobalt transport protein (316 aa), fasta scores: E(): 0, 99.1% identity in 316 aa overlap. Contains Pfam match to entry PF01544 CorA, CorA-like Mg2+ transporter protein. Contains hydrophobic, possible membrane-spanning regions"

/codon\_start=1

/transl\_table=11

/product="putative magnesium and cobalt transport protein"

/protein\_id="CAD09624.1"

/db\_xref="GI:16505746"

/translation="MLSAPOLENNRRLRLVEESQPLVNAWIDLVDPDDERLLOS ELGQSLATRELDITFASARFFEDDDGLIHSEFFEDADDAHNSVAFTIRGRLEF TLIERELPARLYRMRARSQSWDGNAYELLIDLEFKIRQLADEINITYSDLOLSR VIMEHQGDGEYDLSLTAELIDIGMKNVRCILMDLORALFIVYKARLPGGLOARE ILRDIESLPHNESLFOKAVFLMQAANGFCINIDONRLIKESVSVVFLPPTLVAASSV GMEFEPPELTKMFGYPAALIFMLIAGLAVLYLKKRNWL"

/complement(2697..3584)

/gene="corA"

/note="Pfam match to entry PF01544 CorA, CorA-like Mg2+ transporter protein, score 432.90, E-value 2.8e-126"

/complement(3653..3656)

/note="possible RBS"

4004..4007

/note="possible RBS"

4014..4343

/note="possible RBS"

misc\_feature

gene

RBS

RBS

gene

CDS

/gene="HCM1.08"  
4014..4343  
/gene="HCM1.08"  
/note="HCM1.08, hypothetical protein, len: 109 aa; similar to the N-terminal half of SW:YIG\_ECOLI (EMBL:U02122), y1g, Escherichia coli hypothetical protein (254 aa), fasta scores: E(): 0, 97.9% identity in 95 aa overlap"

gene  
complement(4363..4569)  
/gene="HCM1.10c"  
complement(4363..4569)  
/gene="HCM1.10c"  
/note="HCM1.10c, hypothetical protein, len: 68 aa; unknown function"

CDS  
complement(4363..4569)  
/gene="HCM1.10c"  
complement(4363..4569)  
/gene="HCM1.10c"  
/note="HCM1.10c, hypothetical protein, len: 68 aa; unknown function"

Query Match  
Best Local Similarity 96.6%; Pred. No. 2.3e-26;  
Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ACACACTGCGAGAGCGCCAGACGCTGGCCGCAACACAGCTTAAGACAGCTGGCGGAGAG 63  
175127 ACCCAGCTGCGAGAGCGCCAGACGCTGGCCGCAACACAGCTTAAGACAGCTGGCGGAGAG 175186

QY 64 ATGGCCGACTTGGGCGCATGGAACCGTGTCTGTAAGTGTGTGCGCTGCCATGCA 123  
175187 ATGGCCGACTTGGGCGCATGGAACCGTGTCTGTAAGTGTGTGCGCTGCCATGCA 175246

QY 124 CGAAGGGGAATGTTTCTGCGCCGTTGATCGCTGACTACAGGATCTCAGGC 177  
175247 CGAAGGGGAATGTTTCTGCGCCGTTGATCGCTGACTACAGGATCTCAGGC 175300

Db 175247 CGAAGGGGAATGTTTCTGCGCCGTTGATCGCTGACTACAGGATCTCAGGC 175300

RESULT 13  
ARMERGE4 516 bp DNA linear BCT 31-DEC-1994  
LOCUS A.faecalis (SE20MERR) merr gene for regulatory protein.  
DEFINITION  
ACCESSION Z33484  
VERSION Z33484.1 GI:607038  
KEYWORDS merr; regulatory protein.  
SOURCE Alcalligenes faecalis.  
ORGANISM Bacteria; Proteobacteria; beta subdivision; Alcalligenaceae;  
Alcalligenes.  
REFERENCE 1 (bases 1 to 516)  
AUTHORS Osborn, A.M., Bruce, R.D., Strike, P. and Ritchie, D.A.  
TITLE Sequence Conservation Between Regulatory Mercury Resistance Genes  
JOURNAL from Mercury Polluted and Pristine Environments  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 516)  
Osborn, A.M.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-1994) Osborn A. M., University of Liverpool,  
Genetics and Microbiology, Liverpool, Merseyside, UK, L69 3BX  
FEATURES  
source  
1..516  
/organism="Alcalligenes faecalis"  
/isolate="SE20"  
/db\_xref="taxon:511"  
/clone="pSE20R1, pSE20R2"

-35\_signal  
complement(13..18)  
/citation=[1]  
-10\_signal  
complement(28..33)  
/citation=[1]  
-10\_signal  
36..41  
/citation=[1]  
-35\_signal  
61..66  
/citation=[1]  
72..506  
/gene="merr"  
72..506  
/gene="merr"  
/citation=[1]  
/codon\_start=1  
/transl\_table=1  
/product="regulatory protein"  
/protein\_id="CAA8382.1"  
/db\_xref="GI:607039"  
/db\_xref="SPTREMBL:Q44190"  
/translation="MENNLENTTGVFAKAGVNETTFYORKILLPEBDKPYGSIR  
RYGADYRVKFKSAORLGFSLDEIAELRLDGTGHCSEASSLAHKLKDVREKMD  
LARMEYVLSLVCACHARKGNVSCPLIASIAGEGGLARSAMP"

BASE COUNT 112 a 130 c 165 g 109 t

ORIGIN

Query Match  
Best Local Similarity 96.0%; Pred. No. 7.3e-26;  
Matches 167; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 ACACACTGCGAGAGCGCCAGACGCTGGCCGCAACACAGCTTAAGACAGCTGGCGGAGAG 63  
309 ACCCAGCTGCGAGAGCGCCAGACGCTGGCCGCAACACAGCTTAAGACAGCTGGCGGAGAG 368

QY 64 ATGGCCGACTTGGGCGCATGGAACCGTGTCTGTAAGTGTGTGCGCTGCCATGCA 123  
369 ATGGCCGACTTGGGCGCATGGAACCGTGTCTGTAAGTGTGTGCGCTGCCATGCA 428

QY 124 CGAAGGGGAATGTTTCTGCGCCGTTGATCGCTGACTACAGGATCTCAGGC 177  
429 CGAAGGGGAATGTTTCTGCGCCGTTGATCGCTGACTACAGGATCTCAGGC 482

Db 429 CGAAGGGGAATGTTTCTGCGCCGTTGATCGCTGACTACAGGATCTCAGGC 482

RESULT 14  
EMERGE8 516 bp DNA linear BCT 31-DEC-1994  
LOCUS E.aerogenes merr gene for regulatory protein.  
DEFINITION  
ACCESSION Z33488  
VERSION Z33488.1 GI:607071  
KEYWORDS merr; regulatory protein.  
SOURCE Enterobacter aerogenes.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Enterobacter.  
REFERENCE 1 (bases 1 to 516)  
AUTHORS Osborn, A.M., Bruce, R.D., Strike, P. and Ritchie, D.A.  
TITLE Sequence Conservation Between Regulatory Mercury Resistance Genes  
JOURNAL from Mercury Polluted and Pristine Environments  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 516)  
Osborn, A.M.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-1994) Osborn A. M., University of Liverpool,  
Genetics and Microbiology, Liverpool, Merseyside, UK, L69 3BX  
FEATURES  
source  
1..516  
/organism="Enterobacter aerogenes"  
/isolate="T238"  
/db\_xref="taxon:548"  
/clone="pT238R1, pT238R2"  
complement(13..18)  
/citation=[1]  
complement(28..33)  
/citation=[1]



-10\_signal 36.41  
/citation=[1]  
-35\_signal 61.66  
/citation=[1]  
gene 72.506  
/gene="mer"  
72.506  
/gene="mer"  
/gene="mer"  
/citation=[1]  
/codon\_start=1  
/transl\_table=11  
/product="regulatory protein"  
/protein\_id="CAA83896.1"  
/db\_xref="GI:607072"  
/db\_xref="SPTREMBL:046655"  
/translation="MENNLESLTIGAKAGVAVNTIRFYORKGLPEPPKPYGSIR  
RYGADVTVRVRYKSNORIGFSIDEIALLRUDGTHCEBASSLAHKLQDVERKMD  
LARMETVSELVCACHARKGNVSCPLIASLQGGTSLGSAAP"  
BASE COUNT 103 a 138 c 169 g 106 t  
ORIGIN

Query Match 45.2%; Score 145.2; DB 1; Length 516;  
Best Local Similarity 89.7%; Pred. No. 5e-22;  
Matches 156; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 4 ACACACTGCGAGAGGCCAGACCTGGCCGGAACACAGCTCAAGAGCTGCGGAGAG 63  
|||  
DB 309 ACCCACTGCGAGAGGCCAGACCTGGCCGAGACAGCTTCAGACGTGCGGAGAG 368  
|||  
QY 64 ATGGCCGACTGGCGGCGCATGGAACCGTGTCTGTGAAGTGTGCGCTGCGCATGCA 123  
|||||  
DB 369 ATGGCCGACTGGCGGCGCATGGAACCGTGTCTGTGAAGTGTGCGCTGCGCATGCG 428  
|||||  
QY 124 CGAAGGGGGAATGTTCTGCGCTGTGATGCGCTCACTACAGGATCTCAGGC 177  
|||  
DB 429 CGAAGGGGGAATGTTCTGCGCTGTGATGCGCTCACTACAGGATCTCAGGC 482  
|||

RESULT 15  
AF092069/c 333 bp DNA linear BCT 02-SEP-1999  
LOCUS Pseudomonas putida group 2 plasmid merR (mer) gene, partial cds.  
DEFINITION  
ACCESSION AF092069  
VERSION AF092069.1 GI:5817323  
KEYWORDS  
SOURCE Pseudomonas putida.  
ORGANISM Pseudomonas putida  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.

REFERENCE 1 (bases 1 to 333)  
AUTHORS Bruce, K.D., Lilley, A.K. and Bailey, M.J.  
TITLE mer sequences on plasmids  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 333)  
AUTHORS Bruce, K.D., Lilley, A.K. and Bailey, M.J.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1998) School of Biological Sciences, University  
of Liverpool, Crown Street, Liverpool, Merseyside L69 7ZD, UK  
FEATURES  
Source Location/Qualifiers  
1..333

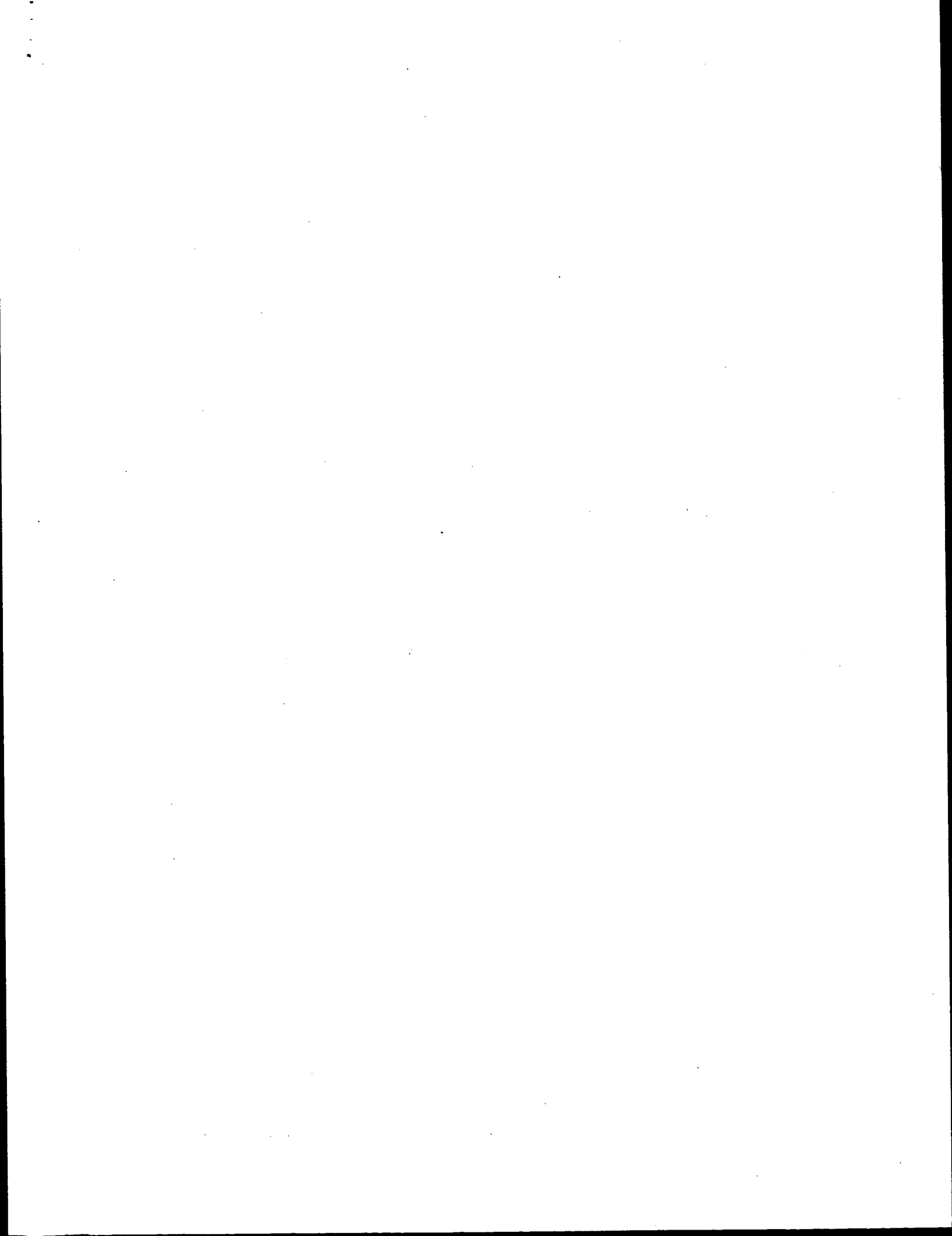
/organism="Pseudomonas putida"  
/strain="KT2440"  
/db\_xref="taxon:303"  
/plasmid="group 2 plasmid"  
/note="this plasmid is naturally-occurring but was  
recovered in a Pseudomonas strain which was introduced  
into the samples"  
complement(<1..>333)  
/gene="mer"  
complement(<1..>333)  
/gene="mer"  
/note="regulatory protein (repressor/inducer)"  
/codon\_start=1

/transl\_table=11  
/product="Mer"  
/protein\_id="A052705.1"  
/db\_xref="GI:5817324"  
/translation="IRFYORKGLVPGPEOPYSIPRYGADVTVRVRYKSNORIGFSL  
DEFAELRLDDGTHCEBASSLAHKLQDVERKMDLARMETVSELVCACHARKGNV  
CPILIESLQA"  
BASE COUNT 67 a 109 c 94 g 63 t  
ORIGIN

Query Match 43.8%; Score 140.6; DB 1; Length 333;  
Best Local Similarity 91.4%; Pred. No. 5.2e-21;  
Matches 149; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 4 ACACACTGCGAGAGGCCAGACCTGGCCGGAACACAGCTCAAGAGCTGCGGAGAG 63  
|||  
DB 165 ACCCACTGCGAGAGGCCAGACCTGGCCGAGACAGCTTCAGACGTGCGGAGAG 106  
|||  
QY 64 ATGGCCGACTGGCGGCGCATGGAACCGTGTCTGTGAAGTGTGCGCTGCGCATGCA 123  
|||||  
DB 105 ATGGCCGACTGGCGGCGCATGGAACCGTGTCTGTGAAGTGTGCGCTGCGCATGCG 46  
|||||  
QY 124 CGAAGGGGGAATGTTCTGCGCTGTGATGCGCTCACTACAGG 166  
|||  
DB 45 CGAAGGGGGAATGTTCTGCGCTGTGATGCGCTCACTACAGG 3  
|||

Search completed: June 2, 2003, 06:52:54  
Job time : 1800 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 04:50:18 ; Search time 199 Seconds  
(Without alignments)  
3632.620 Million cell updates/sec

Title: US-09-977-137A-3

Perfect score: 321  
Sequence: 1 atgacacactgcgagagagc.....agggaatgtcttcgcgcg 321

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	321	100.0	321	24	ABK52207
2	321	100.0	354	24	ABK52212
3	321	100.0	509	24	ABK52213
4	164.4	51.2	435	24	ABK52206
5	164.4	51.2	7076	13	AAQ20186
6	164.4	51.2	7076	16	AAQ86906
7	129.2	40.2	568	15	AAQ85554
8	43.6	13.6	3012	21	AAZ52054
9	43.6	13.6	8908	21	AAZ52055

10	41.2	12.8	7897	24	ABK91609	Modified HIV prote
11	41.2	12.8	9166	24	ABK91616	Modified HIV prote
12	41.2	12.8	9167	24	ABK91619	Modified HIV prote
13	41.2	12.8	9169	24	ABK91617	Modified HIV prote
14	41.2	12.8	9170	24	ABK91614	Modified HIV prote
15	41.2	12.8	9189	24	ABK91607	Modified HIV prote
16	41.2	12.8	9194	24	ABK91611	Modified HIV prote
17	41.2	12.8	9194	24	ABK91612	Modified HIV prote
18	41.2	12.8	9407	24	ABK91623	Modified HIV prote
19	41.2	12.8	9782	24	ABK91621	Modified HIV prote
20	41.2	12.8	9783	24	ABK91620	Modified HIV prote
21	41.2	12.8	9785	24	ABK91624	Modified HIV prote
22	41.2	12.8	9785	24	ABK91618	Modified HIV prote
23	41.2	12.8	12411	24	ABK91613	Modified HIV prote
24	39.6	12.3	9788	24	ABK91622	Modified HIV prote
25	38.4	12.0	4341	24	AA167897	Nucleotide sequenc
26	38.4	12.0	4343	24	AA167896	Nucleotide sequenc
27	37.8	11.8	3519	16	AAQ94352	BHV1 g1 glycoprote
28	36.4	11.3	4319	21	AA470414	HIV Gag-polymerase
29	36.2	11.3	2799	19	AAV49431	Bovine herpesvirus
30	36.2	11.3	3382	13	AAQ29515	Bovine herpesvirus
31	36.2	11.3	3382	14	AAQ48249	Bovine herpesvirus
32	36	11.2	902	23	ABL02087	Drosophila melanog
33	36	11.2	2963	23	ABL02086	Pseudomonas aerugi
34	35.8	11.2	1431	23	AA554175	Gene encoding a su
35	35.8	11.2	10732	21	AA10594	Beta-catenin nucle
36	35	10.9	1933	24	ABK47635	CDNA encoding huma
37	35	10.9	2215	24	ABK47638	Propionibacterium
38	34.8	10.8	24187	23	AA559529	M. capsulatus gene
39	34.2	10.7	378	24	ABQ91799	Mycobacterium tube
40	34.2	10.7	4403765	22	AA199683	HIV-1 integrase ge
41	34.2	10.7	4411529	22	AA199682	Domestic mite Bt11
42	34	10.6	930	21	AA668143	Domestic mite Bt11
43	34	10.6	2598	22	AAQ06235	Domestic mite Bt11
44	34	10.6	2625	22	AAQ06237	Domestic mite Bt11
45	34	10.6	3111	22	AAQ06236	Domestic mite Bt11

## ALIGNMENTS

RESULT 1	ABK52207	standard; DNA; 321 BP.
ID	ABK52207	
XX	ABK52207	
AC	ABK52207	
XX	13-AUG-2002	(first entry)
XX		
DE	DNA encoding synthetic Merr chelon variant.	
XX		
KW	Mercuric ion; contaminated soil; ground water; hydroponic solution;	
KW	irrigation water; waste stream; contaminated aqueous medium;	
KW	biological fluid; gastrointestinal tract; Chelon protein;	
KW	enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;	
KW	heavy metal binding protein; Merr; gene; ds.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..321
FT		/*tag= a
FT		/partial
FT		/product= "Synthetic Merr chelon variant"
FT		/note= "This sequence encodes the first 107 amino acids
FT		of the synthetic Merr chelon variant. This
XX		sequence lacks a stop codon"
XX	MO200230962-A2.	
XX	18-APR-2002.	
PD		
XX	12-OCT-2001; 2001MO-US31819.	







XX 31-MAR-1996 (first entry)  
DT XX  
DE Vibrio cholerae disrupted haemolysin gene.  
XX  
KM Vibrio cholerae: avirulent; serogroup: immunogenic; vaccine:  
XX cholera; toxin: mercury resistance; cholera toxin B subunit: ss.  
OS  
XX Vibrio cholerae.  
XX  
PN W09510300-A1.  
XX  
PD 20-APR-1995.  
XX  
PF 07-OCT-1994; 94MO-US11424.  
XX  
PR 08-OCT-1993; 93US-013438.  
XX 08-OCT-1993; 93US-013439.  
XX  
PA (UYMA-) UNITV MARYLAND BALTIMORE.  
XX  
PI Kaper JB, Levine MM;  
XX WPI; 1995-161574/21.  
DR  
PT New avirulent strains of Vibrio cholerae expressing toxin B subunit  
PT - are stable against reversion and useful in oral vaccines to  
PT induce local protective immunity  
XX  
PS Example 11; Figure 21; 109pp; English.  
XX  
XX New avirulent Vibrio cholerae strains of a non-O1 serogroup can be  
CC used in vaccines to protect against cholera caused by non-O1 strains,  
CC optionally together with CV11 which protects against O1 strains.  
CC The new strains are derived from the O139 serotype and the DNA  
CC encoding mercury resistance and the cholera toxin B subunit are  
CC inserted into a haemolysin gene. Expression of the B subunit ensures  
CC that the new strains are immunogenic, inducing a strong local  
CC immunity and the mercury resistance gene allows the vaccine strain  
CC to be identified without the use of antibiotics. The new strains do  
CC not revert to the cholera toxin positive phenotype. For treatment,  
CC approx. 1 million organisms are taken orally, in aqueous NaHCO3  
CC solution.  
SQ  
Sequence 7076 BP; 1578 A; 1975 C; 1968 G; 1555 T; 0 other;  
Query Match 51.2%; Score 164.4; DB 16; Length 7076;  
Best Local Similarity 96.6%; Pred. No. 6.4e-35;  
Matches 168; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 4 ACACACTGCGAGGAGGCCAGCGCTGGCCGACACACAAGCTCAAGAGAGTGGCGAGAG 63  
Db ACCCACTGCGAGGAGGCCAGCGCTGGCCGACACACAAGCTCAAGAGAGTGGCGAGAG 5663  
QY ATGGCCGACTTGGGGCGGCATGGAACCGTGTCTGTAAGTCTGTGTGGCTGGCATGCA 123  
Db ATGGCCGACTTGGGGCGGCATGGAACCGTGTCTGTAAGTCTGTGTGGCTGGCATGCA 5723  
QY 124 CGAAAGGGGGAATGTTCTCTGCTGCCCTGATGCGCGACACAGGAGATCCTCAGGC 177  
Db 5724 CGAAAGGGGGAATGTTCTCTGCTGCCCTGATGCGCGACACAGGAGATCCTCAGGC 5777  
RESULT 7  
ID AA058554 standard; DNA; 568 BP.  
XX  
XX AA058554;  
XX  
XX 16-SEP-1994 (first entry)  
DT XX  
DE Mercury resistant control gene merR(1).  
XX

```

KW Resistance: mercury, selectable marker; Thiobacillus ferrooxidans;
KW transformation; detection; ss.
XX Thiobacillus ferrooxidans.
XX
XX Key Location/Qualifiers
XX FT -35_signal 23..28
XX FT -10_signal 46..51
XX FT -10_signal 46..51
XX FT RBS 76..79
XX FT RBS 76..79
XX FT CDS 89..568
XX FT CDS 89..568
XX FT /*tag= d
XX FT /product= merR(1) gene product.
XX
XX JP06000083-A.
XX
XX PD 11-JAN-1994.
XX
XX PF 17-JAN-1991; 91JP-0018338.
XX
XX PR 17-JAN-1991; 91JP-0018338.
XX
XX PA (AKIT-) AKITA KEN.
XX PA (DOWA ) DOWA MINING CO LTD.
XX
XX DR WPI: 1994-07731/10.
XX DR P-PADB; AAR49668.
XX
XX PT Mercury resistant control gene merR and shuttle vector - for
XX PT enhanced expression of mercury resistance marker in transformed
XX PT Thiobacillus sp.
XX
XX PS Claim 1; Page 2; 26pp; Japanese.
XX
XX CC The mercury resistance genes can be used as selectable markers when
XX CC used to transform other bacteria.
XX
XX SQ Sequence 568 BP; 130 A; 150 C; 164 G; 124 T; 0 other;

Query Match 40.2%; Score 129.2; DB 15; Length 568;
Best Local Similarity 78.3%; Pred. No. 1.3e-25;
Matches 155; Conservative 0; Mismatches 43; Indels 0; Gaps 0

QY 4 ACACACTGCGAGGAGCCAGCAGCGCTGGCCGGAACACAAAGCTCAAGAGTGGCGAGAG 63
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 326 ACCCACTCGATGAAGCCAGCAGCAGCTTGGCCGAGCACAAGCTCAGAGAGTGGCGAAG 385
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 ATGGCGGACTTGGCGCGCATGGAACCGTGTCTGTGAAGTCTGTGGCGCTGCATGCA 123
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 386 ATGGCGGACTTGGCGCGCATGGAAGCGCGTGTCTGTGAAGTCTGTGGCGCTGCATGCG 445
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 CGAAGGAGGATGTTTCTGCGCCGTTGATCGGCTACACAGAGGATCTCTAGCAGCCAC 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 446 CGAAGGAGGAGGATGTTTCTGCGCCGCTTAATGCGTCACTAAAAAATGCGCAGCGTTCG 505
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 TGCAGAGAGCCAGCAGC 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 506 GGTGACGAGCCACTTGC 523
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
ID AA252054 standard; DNA; 3012 BP.
XX
XX AA252054;
XX
XX AC
XX
XX 18-JUL-2000 (first entry)
XX
XX DE Codon optimised Human immunodeficiency virus pol coding region.
XX
XX HIV; gag; packaging cell line; lentivirus; retroviral vector particle.

```







PD 25-APR-2002.  
XX  
XX 14-AUG-2001; 2001WO-US25721.  
XX  
XX 14-AUG-2000; 2000US-225097P.  
XX PR 14-NOV-2000; 2000US-252115P.  
XX PR 28-MAR-2001; 2001US-279257P.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX (CHAD/) CHADRAABARTI B K.  
XX Nabel GJ, Huang Y;  
XX WPI: 2002-452382/48.  
XX  
XX New modified human immunodeficiency virus Env, Gag, Pol of Nef DNA and  
XX PT its encoded protein, useful as vaccines for genetic or protein  
XX PT immunisation for acquired immunodeficiency syndrome or human  
XX PT immunodeficiency virus infection -  
XX  
XX Disclosure; Page 766-769; 794pp; English.  
XX  
XX The invention relates to a nucleic acid molecule encoding a modified HIV  
XX CC (human immunodeficiency virus) protein. The nucleic acid molecule and its  
XX CC encoded protein (modified HIV Env, Gag, Pol or Nef) are useful for  
XX CC genetic or protein immunisation to a host, respectively. In particular  
XX CC these are useful for ameliorating the symptoms of acquired  
XX CC immunodeficiency syndrome (AIDS) or HIV infection and generating an  
XX CC antibody or cytotoxic T lymphocyte (CTL) response against native HIV Env,  
XX CC Gag, Pol or Nef. Furthermore, the nucleic acids, proteins and subsequent  
XX CC compositions are useful for treating or preventing HIV infections or  
XX CC AIDS. Sequences ABR91449-ABR91624 represent modified HIV protein-encoding  
XX CC plasmid DNA sequences of the invention.  
XX  
SQ Sequence 9166 BP; 2227 A; 2662 C; 2513 G; 1764 T; 0 other;  
  
Query Match 12.8%; Score 41.2; DB 24; Length 9166;  
Best Local Similarity 48.3%; Pred. No. 0.12;  
Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
  
QY 12 CGAGGAGCCGACGCTGGCCGGAACACAGCTCAAGACGTCGGCGGAGAAAGTGGCCGA 71  
DB 5568 CGAGGTGATCCCGCCGAGACCGGCGGAGAGACGCGCTTCTCTGTAAGTGGCCGG 5627  
QY 72 CTTGGCGCGCATGGAACCGTGTCTGTAACCTCGTGTGCGCTCCATGACAGAAAGG 131  
DB 5628 CCGCTGGCCCGTGAAGACCGTGTCAACACGACGACCACTTCAACGACACCGCT 5687  
QY 132 GAATGTTTCTGCGCCGTTGATCGCTCACTACAGGATCCTCAAGCAACCACTGCGAGA 191  
DB 5688 GAAAGCGCGCTGTGTGGTGGCGCATCAAGCAGAGTTTGGCATCCCTTCAACCCCCA 5747  
QY 192 GGCACGACCGCTGGCGGAAACCAAGCTCAAGACGTCGGCGGAGAAAGTGGCCGACTG 249  
DB 5748 GAGCCAGGCGGTGATGAGAGCATGAACAGAGCTGAAGAAAGATCATCGGCCAGGTG 5805

## RESULT 12

ABR91619 standard; DNA; 9167 BP.

ID ABR91619;  
AC ABR91619;  
XX  
XX 14-AUG-2002 (first entry)  
XX  
XX Modified HIV protein-encoding plasmid DNA #171.  
XX  
XX HIV; human immunodeficiency virus; gene; ds; circular; cyclic; anti-HIV;  
XX KM Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;  
XX KM acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.  
XX  
XX Synthetic.  
XX OS

PN WO200232943-A2.  
XX  
XX 25-APR-2002.  
XX  
XX 14-AUG-2001; 2001WO-US25721.  
XX  
XX 14-AUG-2000; 2000US-225097P.  
XX PR 14-NOV-2000; 2000US-252115P.  
XX PR 28-MAR-2001; 2001US-279257P.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX (CHAD/) CHADRAABARTI B K.  
XX Nabel GJ, Huang Y;  
XX WPI: 2002-452382/48.  
XX  
XX New modified human immunodeficiency virus Env, Gag, Pol of Nef DNA and  
XX PT its encoded protein, useful as vaccines for genetic or protein  
XX PT immunisation for acquired immunodeficiency syndrome or human  
XX PT immunodeficiency virus infection -  
XX  
XX Disclosure; Page 775-778; 794pp; English.  
XX  
XX The invention relates to a nucleic acid molecule encoding a modified HIV  
XX CC (human immunodeficiency virus) protein. The nucleic acid molecule and its  
XX CC encoded protein (modified HIV Env, Gag, Pol or Nef) are useful for  
XX CC genetic or protein immunisation to a host, respectively. In particular  
XX CC these are useful for ameliorating the symptoms of acquired  
XX CC immunodeficiency syndrome (AIDS) or HIV infection and generating an  
XX CC antibody or cytotoxic T lymphocyte (CTL) response against native HIV Env,  
XX CC Gag, Pol or Nef. Furthermore, the nucleic acids, proteins and subsequent  
XX CC compositions are useful for treating or preventing HIV infections or  
XX CC AIDS. Sequences ABR91449-ABR91624 represent modified HIV protein-encoding  
XX CC plasmid DNA sequences of the invention.  
XX  
SQ Sequence 9167 BP; 2225 A; 2668 C; 2517 G; 1757 T; 0 other;  
  
Query Match 12.8%; Score 41.2; DB 24; Length 9167;  
Best Local Similarity 48.3%; Pred. No. 0.12;  
Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
  
QY 12 CGAGGAGCCGACGCTGGCCGGAACACAGCTCAAGACGTCGGCGGAGAAAGTGGCCGA 71  
DB 5569 CGAGGTGATCCCGCCGAGACCGGCGGAGAGACCGCTTCTCTGTAAGTGGCCGG 5628  
QY 72 CTTGGCGCGCATGGAACCGTGTCTGTAACCTCGTGTGCGCTCCATGACAGAAAGG 131  
DB 5629 CCGCTGGCCCGTGAAGACCGTGTCAACACGACGACCACTTCAACGACACCGCT 5688  
QY 132 GAATGTTTCTGCGCCGTTGATCGCTCACTACAGGATCCTCAAGCAACCACTGCGAGA 191  
DB 5689 GAAAGCGCGCTGTGTGGTGGCGCATCAAGCAGAGTTTGGCATCCCTTCAACCCCCA 5748  
QY 192 GGCACGACCGCTGGCGGAAACCAAGCTCAAGACGTCGGCGGAGAAAGTGGCCGACTG 249  
DB 5749 GAGCCAGGCGGTGATGAGAGCATGAACAAAGAGCTGAAGAAAGATCATCGGCCAGGTG 5806

## RESULT 13

ABR91617 standard; DNA; 9169 BP.

ID ABR91617;  
AC ABR91617;  
XX  
XX 14-AUG-2002 (first entry)  
XX  
XX Modified HIV protein-encoding plasmid DNA #169.  
XX  
XX HIV; human immunodeficiency virus; gene; ds; circular; cyclic; anti-HIV;  
XX KM Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;  
XX KM acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.  
XX  
XX Synthetic.  
XX OS



KW HIV; human immunodeficiency virus; gene; ds; circular; cyclic; anti-HIV-  
KW Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;  
KW acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.

OS Synthetic.

PN WO200232943-A2

PD 25-APR-2002.

PF 14-AUG-2001; 2001MO-US25721

PR	14-AUG-2000; 2000US-22509/P
PR	14-NOV-2000; 2000US-252115P

PR 28-MAR-2001; 200105-2/925/P  
XX  
XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES  
PA (CHAD/) CHADPARTI B K.

XX Nabe] GT. Huang Y:  
PT

XX  
DB  
WPT: 2002-452382/48.

New modified human immunode

PT immunisation for acquired

XX  
XX

XX

The invention relates to a nucleic acid molecule encoding a modified HIV (human immunodeficiency virus) protein. The nucleic acid molecule and its encoded protein (modified HIV Env, Gag, Pol or Nef) are useful for genetic or protein immunisation to a host, respectively. In particular these are useful for ameliorating the symptoms of acquired immunodeficiency syndrome (AIDS) or HIV infection and generating an antibody or cytotoxic T lymphocyte (CTL) response against native HIV Env, Gag, Pol or Nef. Furthermore, the nucleic acids, proteins and subsequent compositions are useful for treating or preventing HIV infections or AIDS. Sequences ABR91449-ABR91624 represent modified HIV protein-encoding plasmid DNA sequences of the invention.

SQ Sequence 9189 BP; 2232 A; 2

Query Match	12.887
PostgreSQL 9.4.12	48.387

Matches 115; Conservative

QY 12 CGAGGAGGCCAGGAGCCT

Db 5572 CGAGGTGATCCCGCG

1 2 CTTGGCGCGCATGGAAAC  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

DB 5632 CCGTGGCCCGTGAAGAA

132 6AA1G111CC1GCCCC1  
QY

LD 0032 GAAAGCCCCCAGCCACGCG

192 GGCACGCGCCAGCCTGGCCTG  
GAT

[illegible]

Search completed: June 2, 2003

000000 : 000000

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 05:02:13 ; Search time 1515 seconds  
(without alignments)  
3431.519 Million cell updates/sec

Title: US-09-977-137A-3

Perfect score: 321

Sequence: 1 atgacacactgcgagagagc.....aggggatgttctcgcgcg 321

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlin:\*  
5: em\_estlin:\*  
6: em\_estlin:\*  
7: em\_estlin:\*  
8: em\_estlin:\*  
9: em\_estlin:\*  
10: em\_estlin:\*  
11: em\_estlin:\*  
12: em\_estlin:\*  
13: em\_estlin:\*  
14: em\_estlin:\*  
15: em\_estlin:\*  
16: em\_estlin:\*  
17: em\_estlin:\*  
18: em\_estlin:\*  
19: em\_estlin:\*  
20: em\_estlin:\*  
21: em\_estlin:\*  
22: em\_estlin:\*  
23: em\_estlin:\*  
24: em\_estlin:\*  
25: em\_estlin:\*  
26: em\_estlin:\*  
27: em\_estlin:\*

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	40.6	12.6	1101	CNS017SY	AL108460 Drosophila
2	39	12.1	10	BE402139	BE402139 CSB004H01
3	39	12.1	587	BQ607497	BQ607497 BRY_3391
4	38.2	11.9	524	BI341447	BI341447 368872 MA
5	38	11.8	736	BH179075	BH179075 013.D_05-
6	38	11.8	736	CNS07LE8T	AL616031 T3 end of

7	37.8	11.8	716	10	BE636656	BE636656 rockefell
8	37	11.5	710	10	BE601964	BE601964 HVSMH010
9	36.8	11.5	479	12	BF484898	BF484898 WHE2320.H
10	36.8	11.5	538	14	BO805185	BO805185 WHE3563.G
11	36.6	11.5	640	10	AV651765	AV651765 AV651765
12	36.6	11.4	304	10	AM177874	AM177874 IL3-HT005
13	36.6	11.4	352	10	AV627687	AV627687 AV627687
14	36.6	11.4	707	12	BE6857743	BE6857743 102405360
15	36.4	11.3	480	13	BM488564	BM488564 psm2n.pk0
16	36.4	11.3	657	13	BM635756	BM635756 170006875
17	36.4	11.3	663	13	BM625043	BM625043 170006874
18	36.4	11.3	669	13	BM656350	BM656350 170006873
19	36.4	11.3	689	13	BM619285	BM619285 170006873
20	36.4	11.3	696	12	BE497931	BE497931 602543255
21	36.4	11.3	699	13	BM636151	BM636151 170006875
22	36.4	11.3	707	13	BM580657	BM580657 170006872
23	36.2	11.3	302	14	BO812467	BO812467 1030030C1
24	36.2	11.3	623	10	BE583640	BE583640 8-11B-HA
25	36	11.2	461	10	AV390467	AV390467 AV390467
26	36	11.2	479	10	AV632072	AV632072 AV632072
27	36	11.2	527	10	AV392278	AV392278 AV392278
28	36	11.2	528	10	AV392451	AV392451 AV392451
29	35.8	11.2	859	12	BE627594	BE627594 HVSMH000
30	35.6	11.1	290	10	AM807446	AM807446 MR4-ST006
31	35.6	11.1	456	10	AM707773	AM707773 832012P07
32	35.6	11.1	485	10	AM707771	AM707771 832012P06
33	35.4	11.0	299	10	AM807414	AM807414 MR4-ST006
34	35.4	11.0	413	10	AM462578	AM462578 BP230009B
35	35.4	11.0	415	13	BM369560	BM369560 EBem07-SO
36	35.4	11.0	487	14	BO159851	BO159851 WHE2237.D
37	35.4	11.0	502	14	BO159422	BO159422 WHE2204.H
38	35.4	11.0	514	12	BE292042	BE292042 WHE2203.H
39	35.4	11.0	518	9	AI989014	AI989014 fs29506.Y
40	35.4	11.0	540	10	AV704072	AV704072 AV704072
41	35.4	11.0	582	13	BM525809	BM525809 sak72b07
42	35.4	11.0	764	14	BO840689	BO840689 WHE4202.D
43	35.2	11.0	491	13	BI541598	BI541598 455406 MA
44	35.2	11.0	506	14	BO528756	BO528756 3524.1.40
45	35.2	11.0	515	10	AV602443	AV602443 AV602443

## ALIGNMENTS

RESULT 1  
CNS017SY 1101 bp DNA linear GSS 26-JUL-1999  
LOCUS  
DEFINITION  
BACN37L08 of Drosophila genome survey sequence SP6 end of BAC  
BACN37L08 of Drosophila library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
AL108460  
VERSION  
GI:5628764  
SOURCE  
Drosophila melanogaster.

ORGANISM  
Drosophila melanogaster.

REFERENCE  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.  
Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequences :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

Location/Qualifiers

```

source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37108"
/clone_1lb="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"
BASE COUNT      254 a      176 c      160 g      152 t      359 others
ORIGIN
Query Match      12.6% Score 40.6; DB 17; Length 1101;
Best Local Similarity 13.3%; Pred. No. 1.7;
Matches 35; Conservative 131; Mismatches 97; Indels 0; Gaps 0;

QY 6 ACACGTGGAGAGAGCCAGCAGCTGGCGGAACACAGCTCAAGACGTGGCGGAGAGAT 65
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 782 ACSASCSASCMGCVSSGSSCSASGSCGCVSSCAVSASASVSMKVASMAVASCVA 841
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 66 GCGCGACTTGGCGCGCAGAAACCGTCTGTCTGACCTGCTGTGGCGCTGCATGCACG 125
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 842 SGMAGAVSSSCRSVSVASVSAASVSSSSSVSAASVSAASVSSSSSVSSASMAVAA 901
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 126 AAGGGAGATGTTCTGCGCCGCTGATGCGCTACACAGGATCCTCAGGCGACCCACTG 185
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 902 AAAYAVSVSVASVSVSSSSSSSSASVSVSVSAASASVSSSSSVSTSSASVSVS 961
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 186 CGAGAGAGCCAGCAGCTGGCGGACACAGCTCAAGACGTGGCGGAGAGATGGCCGA 245
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 962 VSAVSMANAVSSASSSSSSVSVSVVAAVAAAAAASSSSSASASVAVSVSSSS 1021
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 246 CTTGGCGCGCATGGAACCGTGC 268
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1022 SSASSSSSSSASSSSVSVSSSC 1044

RESULT 2
BE402139      587 bp      mRNA      linear      EST 21-JUN-2000
LOCUS      CSB004H01F990908 ITTC CSB wheat Endosperm Library Triticum aestivum
DEFINITION      cDNA clone CSB004H01, mRNA sequence.
ACCESSION      BE402139
VERSION      BE402139.1 GI:9361607
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticaceae; Triticum.
REFERENCE      1 (bases 1 to 587)
AUTHORS      Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pechonin, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorells, M., Warburton, M. and Wenzel, G.
JOURNAL      International Triticaceae EST Cooperative (ITTC): Production of
COMMENT      Expressed Sequence Tags for Species of the Triticaceae
Unpublished (2000)
CONTACT      Appels R
Div. of Plant Industry, CSIRO
Canberra ACT 2601 AUSTRALIA
Tel: 61 62 465496
Fax: 61 62 465000
Email: rudiepl.csiro.au
International Triticaceae EST Cooperative (ITTC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. .587
/organism="Triticum aestivum"
/cultivar="WYuna"
/db_xref="taxon:4565"
/clone="CSB004H01"
/clone_1lb="ITTC CSB wheat Endosperm Library"

FEATURES
source
1. .587
/organism="Triticum aestivum"
/cultivar="WYuna"
/db_xref="taxon:4565"
/clone="CSB004H01"
/clone_1lb="ITTC CSB wheat Endosperm Library"

```

```

/tissue_type="endosperm"
/dev_stage="8-12 days post anthesis"
/lab_host="Escherichia coli SGR"
/note="Vector: Lambda Zap/Bluescript; Site 1: XhoI;
Site 2: EcoRI; Plants grown in Phytochthon with 18C/13C
(day/night) 16 hour light. M13 Reverse sequencing primer
used. 1.0 Kbp average insert size."
BASE COUNT      113 a      218 c      185 g      71 t
ORIGIN
Query Match      12.1% Score 39; DB 10; Length 587;
Best Local Similarity 47.1%; Pred. No. 3.5;
Matches 120; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 21 CAGCAGCCTGGCGGCAACACAGCTCAAGGACGCGCGAGAGATGGCGCGTGGCGCG 80
   || || || || || || || || || || || || || || || || || || || || || ||
DB 272 CGGCGAGAGAGCGCGGAGCGAGCGCCGACGAGCGCGAGCGGAGAGAGACGAGCGCGT 331
   || || || || || || || || || || || || || || || || || || || || || ||
QY 81 CATGGAACCCGCTGCTGTCTGAACCTGCTGCGCGCTGCCATGACGAAAGGGAATGTTTC 140
   || || || || || || || || || || || || || || || || || || || || || ||
DB 332 GATGAGAGTGCAGCGCGCGCGGAGAGAGAGTGCAGAGAGTGCAGCGCGCAAGCGCGC 391
   || || || || || || || || || || || || || || || || || || || || || ||
QY 141 CTGCGCGTGTATGCGCTGCTACTAGAGGATCTTCAAGCACCCACTGCGAGAGAGCGCAGAG 200
   || || || || || || || || || || || || || || || || || || || || || ||
DB 392 CGCCAAAGGCGCTCGGAGCAAGAGCAGCTCGCGCGCGCGCTGCGCACGTTTCCGACGG 451
   || || || || || || || || || || || || || || || || || || || || || ||
QY 201 CTTGGCGGAAACAGCTCAAGAGCGTGGCGGAGAGATGGCGACTTGGCGCGCATGA 260
   || || || || || || || || || || || || || || || || || || || || || ||
DB 452 GCTCGCGAGCTCAACATGAGCGCTTACGACGAGAGAGAGAGAGCGGCTCGAGTTTTCAG 511
   || || || || || || || || || || || || || || || || || || || || || ||
QY 261 AACGCTGCTGTCTGA 275
   || || || || || || || || || || || || || || || || || || || || || ||
DB 512 CACTGGTTGGGGGA 526

RESULT 3
BO607497      587 bp      mRNA      linear      EST 25-JUN-2002
LOCUS      BRY.3391 wheat EST endosperm library Triticum aestivum cDNA 5',
DEFINITION      mRNA sequence.
ACCESSION      BO607497
VERSION      BO607497.1 GI:21556826
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticaceae; Triticum.
REFERENCE      1 (bases 1 to 587)
AUTHORS      Clarke, B., Lambrecht, M. and Rhee, S.
JOURNAL      Assessing the utility of Arabidopsis genomic information for
COMMENT      interpreting wheat EST sequences
Unpublished (2002)
CONTACT      Lambrecht M
The Arabidopsis Information Resource
Carnegie Institution of Washington, Dept. of Plant Biology
260 Panama Street, Stanford, CA 94305, USA
Tel: 1 650 325 1521 x 251
Fax: 1 650 325 3748
Email: rhees@coma.stanford.edu.
Location/Qualifiers
1. .587
/organism="Triticum aestivum"
/cultivar="WYuna"
/db_xref="taxon:4565"
/clone_1lb="wheat EST endosperm library"
/tissue_type="endosperm"
/dev_stage="developing endosperm tissue 8, 10 and 12 DPA
(days post anthesis)"
/note="Vector: Bluescript II SK(-)"
BASE COUNT      113 a      218 c      185 g      71 t
ORIGIN

```

Query Match 12.1%; Score 39; DB 14; Length 587;  
 Best Local Similarity 47.1%; Pred. No. 3.5;  
 Matches 120; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 21 CAGCAGCGGCGGAGACACACAGCTCAAGAGCTCGCGAGAGATGGCCGACTTGGCGCG 80  
 Db 272 CGGCAAGAGAGGCGGAGCGACCGCGGAGCGCGGAGCGAGCGAGCGAGCGAGCGCGT 331  
 QY 81 CATGGAACCGTCTGTCACTGCTGTGAGTGGCGCTGCATGACAGAAAGGGAGTGTTC 140  
 Db 332 GATGAGAGTTCAGAGCGGCGGAGAGAGAGTTCAGAGTTCGCGCGCGAGCGCGC 391  
 QY 141 CTGCGCGTGTGATGCTGCTCACTACAGGAGTCTCAGGACCGCCTGCGAGAGCGAGCAG 200  
 Db 392 CGCAGAGGCGCTGCGAGCAGAGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 451  
 QY 201 CCTGGCGGAGACAGCTCAAGAGAGCGCGGAGAGAGATGGCGCGCTGGCGCGCGCG 260  
 Db 452 GCTCGCGAGCTCACTACAGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511  
 QY 261 AACCTGCTGTCTGA 275  
 Db 512 CACTGCTGTGGGGA 526

RESULT 4  
 LOCUS B1341447 524 bp mRNA linear EST 30-JUL-2001  
 DEFINITION 368872 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.  
 ACCESSION B1341447  
 VERSION B1341447.1 GI:15034736  
 KEYWORDS EST.  
 SOURCE Pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 524)  
 Fahrrekrug, S.C., Freiling, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,  
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,  
 and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Smith TPJ  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel.: 402 762 4366  
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt\_trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.

PCR primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCACAGTACGACG  
 Plate: 108 row: H column: 15  
 Seq primer: ATTGAGGACACTATAG.

FEATURES  
 Location/Qualifiers

source  
 1..524  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_id="MARC 2P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from testis, ovary,  
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 84 a 180 c 165 g 95 t  
 ORIGIN  
 Query Match 11.9%; Score 38.2; DB 13; Length 524;  
 Best Local Similarity 59.8%; Pred. No. 5.5;  
 Matches 64; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 11 GCGAGAGCGCCAGCAGCTGGCGGACACAGCTCAAGAGAGCTGCGGAGAGATGGCGCG 70  
 Db 199 GCGAGCAGGCGCCGAGCTCTGCTGAGCGGAGAGTGAAGAGCTCGGAGAGCTTGGCCG 140  
 QY 71 ACTTGCGCGGAGTGAAGAGCGTCTGTCTGTCACTGCTGTGCGCTGC 117  
 Db 139 CCTGGGCGCAGCAGAACCACTCCGCTGAGGCGCCCGCGCGCGCG 93

RESULT 5  
 LOCUS B179075 736 bp DNA linear GSS 19-OCT-2001  
 DEFINITION 013.D-05-rev SMBAC1 Schistosoma mansoni genomic clone 013D05 5',  
 DNA sequence.  
 ACCESSION B179075  
 VERSION B179075.1 GI:16279543  
 KEYWORDS GSS.  
 SOURCE Schistosoma mansoni.  
 ORGANISM Schistosoma mansoni.

REFERENCE Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigeiida; Schistosomatidae; Schistosomatidae; Schistosoma.  
 1 (bases 1 to 736)  
 Le Paslier, M.-C., Pierce, R.J., Meriin, F., Hirai, H., Wu, W., Williams,  
 D.L., Johnston, D., Llovera, P.T. and Le Paslier, D.  
 Construction and characterization of a Schistosoma mansoni  
 bacterial artificial chromosome library  
 Genomics 65 (2), 87-94 (2000)  
 20247247  
 Other\_GSSs: 013.D-05-21  
 Contact: Pierce RJ  
 INSERM U 167

TITLE Institut Pasteur de Lille  
 1 rue du Professeur A. Calmette, 59019-Lille, France  
 Tel.: (33) (0)3 20877783  
 Fax: (33) (0)3 20877888  
 Email: Raymond.Pierce@pasteur-lille.fr  
 CNS sequencing ID=DC0AA013CB03BP1  
 Plate: 013 row: D column: 05  
 Seq primer: M13 reverse primer  
 Class: BAC ends  
 High quality sequence stop: 736.

FEATURES  
 Location/Qualifiers

source  
 1..736  
 /organism="Schistosoma mansoni"  
 /strain="Puerto-Rican"  
 /db\_xref="taxon:6183"  
 /clone\_id="013D05"  
 /clone\_id="SMBAC1"  
 /sex="mixed"  
 /dev\_stage="cercariae"  
 /lab\_host="Biomphalaria glabrata"  
 /note="Vector: pReloBAC 11; Site\_1: Hind III; Partially  
 Hind III digested and size-selected S. mansoni cercarial  
 DNA was ligated into Hind III digested pReloBAC 11 vector  
 and used to transform E. coli DH10B. The complete library  
 contains 23808 clones from 4 independent  
 sizing-ligation-transformations. Average insert size  
 ranges from 70-127 kb and genome coverage is 7.9-fold."

BASE COUNT 132 a 253 c 247 g 99 t  
 ORIGIN  
 Query Match 11.8%; Score 38; DB 17; Length 736;  
 Best Local Similarity 55.5%; Pred. No. 7;  
 Matches 71; Conservative 1; Mismatches 56; Indels 0; Gaps 0;

QY 179 CCGACTGCGAGAGCGCCAGCAGCTGGCGGACACAGCTCAAGAGAGCTGCGGAGAGA 238  
 Db 494 CCAAGTCCAGAGAGGTCAGATTTCAGCGCCCTGAGCAGAGAGAGAGCGAGTTCAGAGAAGC 553  
 QY 239 TGGCGCACTGGGCGGCGGAGAGAGCGTGTGTCTGACATCGTGTGCGCGCTGCATGCAC 298  
 Db 554 AGGCCAAGCATGCTGTGGGCGGAGAGAGCTATTTTCGCGAGCGCTGTGCGCM 613

QY 299 GAAAGGCG 306  
 Db 614 GAAAGGCG 621  
 RESULT 6  
 LOCUS CNS0718T  
 DEFINITION T3 end of clone 013CB03 of library SmbAC1 from strain Puerto-Rican  
 ACCESSION AL616031 GI:16029255  
 VERSION AL616031  
 KEYWORDS Schistosoma mansoni.  
 SOURCE Schistosoma mansoni.  
 ORGANISM Schistosoma mansoni.  
 REFERENCE 1 (bases 1 to 736)  
 AUTHORS Le Paslier, M.C., Pierce, R.J., Meriin, F., Hirai, H., Wu, W., Williams, D., Johnston, D., Loverde, P.T. and Le Paslier, D.  
 TITLE Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library  
 JOURNAL Genomics 65 (2), 87-94 (2000)  
 MEDLINE 2024/247  
 PUBMED 10783255  
 REFERENCE 2 (bases 1 to 736)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage ; BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
 COMMENT Partially Hind III digested and size-selected S. mansoni cercarial DNA was ligated into Hind III digested pRelOBAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold.  
 FEATURES  
 source  
 1..736  
 /organism="Schistosoma mansoni"  
 /strain="Puerto-Rican"  
 /db\_xref="taxon:6183"  
 /clone="013CB03"  
 /clone\_11b="SmbAC1"  
 /note="end : T3"  
 BASE COUNT 132 a 253 c 247 g 99 t 5 others  
 ORIGIN  
 Query Match 11.8% Score 38; DB 17; Length 736;  
 Best Local Similarity 55.5% Pred. No. 7;  
 Matches 71; Conservative 1; Mismatches 56; Indels 0; Gaps 0;  
 QY 179 CCCACTGGAGGAGGCGGAGGCGGCGGAGCAAGCTCAAGAGCGGCGGAGAGA 238  
 Db 494 CCAAGTTCACAAAGGTCAGGTCAGCGCCCTGGACAAAGGCGGAGTTCAGAGAAGC 553  
 QY 239 TGGCCGACTTGGCGCGCATGGAACCGCTGTCTGAATCTGTGTGGCGCTCCATGCAC 298  
 Db 554 AGGCCAAGATGGCGTGGGTCGACATGCTGACACTATTTCCGACAGCGCTGGCTGCGM 613  
 QY 299 GAAAGGCG 306  
 Db 614 GAAAGGCG 621  
 RESULT 7  
 LOCUS BE636656  
 DEFINITION 716 bp mRNA linear EST 03-JAN-2002  
 MASTIGAMOEBALALAMUTHI LAMBDA ZAP II LIBRARY  
 MASTIGAMOEBALALAMUTHI CDNA SIMILAR TO SMALL HEAT SHOCK PROTEIN,  
 mRNA SEQUENCE.  
 ACCESSION BE636656  
 VERSION BE636656.1 GI:9919767

KEYWORDS EST.  
 SOURCE MASTIGAMOEBALALAMUTHI.  
 ORGANISM MASTIGAMOEBALALAMUTHI  
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
 REFERENCE 1 (bases 1 to 716)  
 AUTHORS Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durfee, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.  
 TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostellium, Entamoeba and Mastigamoeba  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 MEDLINE 21819461  
 COMMENT Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 716 Std Error: 0.00  
 POLYA=Yes.  
 FEATURES  
 source  
 1..716  
 /organism="Mastigamoeba balamuthi"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_11b="Mastigamoeba balamuthi lambda ZAP II library"  
 /note="syn: Phreatamoeba balamuthi"  
 BASE COUNT 170 a 228 c 214 g 104 t  
 ORIGIN  
 Query Match 11.8% Score 37.8; DB 10; Length 716;  
 Best Local Similarity 47.6% Pred. No. 7.8; Indels 0; Gaps 0;  
 Matches 111; Conservative 0; Mismatches 122;  
 QY 28 CTGGCGGACACAGCTCAAGAGCTGCGCGAGAGATGCGCCGATGGCGCGATGAA 87  
 Db 277 CTGGACACAGACAAAGTGTGCGCGTGGCGGGGAGAAAGCGCGACGAGAGGCTAC 336  
 QY 88 ACCGTGCTGTGTAAGTCTGTGCGCGCTGCGATGACAGCAAGGAGATGTTCTGCGCG 147  
 Db 337 ACCGCTGTGTAAGTCTGTGCGCGCTGCGATGACAGCAAGGAGATGTTCTGCGCG 396  
 QY 148 TTGATCGCTCACTACAGGAGATCTCAGGACCCAGGCGAGAGAGGCGAGGCTGCGC 207  
 Db 397 GTGACACAGTCTGCGCGTGGCGCGTGGAGGAGCAAGGCGGCTGACGCTGCTGCC 456  
 QY 208 GAACACAGCTCAAGAGCTGCGCGAGAGATGCGCGCGATGTTGCGCGCATGGA 260  
 Db 457 AAGAGCAAGCTGCTGCTCCACAGAAATTTGCTCACTGAGAGCTCACCGA 509  
 RESULT 8  
 LOCUS BE601964  
 DEFINITION 710 bp mRNA linear EST 22-OCT-2001  
 HVSMBH0100J10f Hordeum vulgare 5-45 DAP spike EST library  
 HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMBH0100J10f,  
 mRNA sequence.  
 ACCESSION BE601964  
 VERSION BE601964.2 GI:13190603  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare.  
 ORGANISM Hordeum vulgare.  
 REFERENCE 1 (bases 1 to 710)  
 AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Close, S.J., Oates, R. and Main, D.  
 TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex 5-45 DAP spike cDNA library  
 JOURNAL Unpublished (2001)  
 COMMENT On Aug 21, 2000 this sequence version replaced gi:9859525.  
 Contact: Wing RA



Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hg bases = 229  
Seq primer: AATTAACTCCTCACTAAGG  
High quality sequence start: 6  
High quality sequence stop: 666.

## FEATURES

Source

1. 710  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMEH010010f"  
/clone.lib="Hordeum vulgare 5-45 DAP spike EST library  
HVCNMA0009 (5 to 45 DAP)"  
/tissue\_type="5-45 DAP Spike"  
/lab\_host="SOLR"  
/note="Vector: LambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
plants were grown in the greenhouse at the University of  
California, Riverside (Fenton, SJ Close, TJ Close). Whole  
spikes with awns trimmed were collected at 5, 10, 15, 20,  
30 and 45 DAP (Fenton). Total RNA was prepared from each  
pool, equal quantities of all six RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give phagescript SK(-) cDNA phagmids  
(Choi) in the TJ Close lab at the University of California,  
Riverside. Phagmids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"

BASE COUNT 156 a 209 c 238 g 106 t  
ORIGIN

Query Match 11.5%; Score 37; DB 10; Length 710;  
Best Local Similarity 54.9%; Pred. No. 13;  
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 161 TCAGGAGTCCTCAGCACCCTGCGAGAGAGCCAGCAGCTGGCCGACACAGCTCA 220  
DB 36 TAGAATAACACAGGAGCTCGAGAGATGGGACACAGCTCGAGCGGCGACGCTCA 95  
QY 221 AGAGCTGGCGGAGAGATGGCGAGCTGGCGGAGATGGAACCTGCTGTCTGAATCG 280  
DB 96 AGGAGAGAGTGGGAGACATGAGACCGGCGCGCTCTCTGACCTGGGACCGCTCTCA 155  
QY 281 TGTGGCGCTGCCA 293  
DB 156 ACCGCGCTGGCGA 168

RESULT 9  
BF484898 479 bp mRNA linear EST 06-DEC-2000  
LOCUS  
DEFINITION BFA84898  
ACCESSION  
VERSION BFA84898.1 GI:11568199  
KEYWORDS

## SOURCE

ORGANISM

Bread wheat.  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
; Triticeae; Triticum.  
1 (bases 1 to 479)

## AUTHORS

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han  
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,  
Seaton, C.L. and Tong, J.C.  
The structure and function of the expressed portion of the wheat  
genomes: Pre-anthesis spike cDNA library  
Unpublished (2000)

## TITLE

JOURNAL  
COMMENT  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818

## FEATURES

Source

Email: oanderson@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.  
Location/Qualifiers

1. 479  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE2320.H06.P12"  
/clone.lib="Wheat pre-anthesis spike cDNA library"  
/tissue\_type="Spike before anthesis"  
/dev\_stage="Adult plant"  
/lab\_host="E. coli SOLR"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
greenhouse. Whole spike with awns trimmed, white, green  
and yellow anther were collected and total RNA, and  
poly(A) RNA were prepared, a cDNA library was made, and  
phagmids were in vivo excised to give phagescript  
phagmids in the TJ Close lab (Choi, Close, Fenton) at  
the University of California, Riverside. Plasmid DNA  
preparations and DNA sequencing were performed in the OD  
Anderson lab (all other authors)."

BASE COUNT 81 a 161 c 180 g 57 t  
ORIGIN

Query Match 11.5%; Score 36.8; DB 12; Length 479;  
Best Local Similarity 47.8%; Pred. No. 12;  
Matches 107; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 11 GCGAGAGGCGCAGCAGCTGCGCAACAGCTCAAGAGTGGGAGAGATGGCG 70  
DB 159 GCCAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCT 218  
QY 71 ACTTGGCGGCGATGGAACCGCTCTCTGTAACCTGTGCGCTGCGCATGACAGAAAG 130  
DB 219 CCTCCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCT 278  
QY 131 GGAATGTTCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 190  
DB 279 AGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 338  
QY 191 AGGCGAGAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 234  
DB 339 AGCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 382

RESULT 10  
BQ805185 538 bp mRNA linear EST 31-JUL-2002  
LOCUS  
DEFINITION BQ805185  
ACCESSION  
VERSION BQ805185.1 GI:22029394  
KEYWORDS

KEYWORDS	EST.
SOURCE	bread wheat
ORGANISM	Triticum aestivum
REFERENCE	Eurariophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Triticum.
ATTNORS	1 (bases 1 to 538)
TITLE	Altenbach, S., Anderson, O.D., Chao, S., Chin, A., Close, T.J., Cronin, K., Crossman, C., Fenton, R.D., Lazo, G.R., Pham, J., Rausch, C.J., Wilson, C. and Woo, J.
JOURNAL	The structure and function of the expressed portion of the wheat genomes - Developing grains cdna library
COMMENT	Unpublished (2002) Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel.: 510535773 Fax: 51053595818 Email: oanderson@pw.usda.gov Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: SK primer.
FEATURES	Location/Qualifiers
source	1..538
	/organism="Triticum aestivum"
	/cultivar="Butte 86"
	/db_xref="taxon:4565"
	/clone="WHE3563.G11.M21"
	/clone_1lb="Wheat developing grains cdna library"
	/tissue_type="whole grains"
	/dev_stage="3-44 days post anthesis seed"
	/lab_host="E. coli SOLR"
	/note="Vector: Lambda ZAP II, excised phagemid; Site: 1. EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 24oc/17oc day/night, well-watered, with post-anthesis fertilizer, Environment 2) 24oc/17oc day/night, well-watered, without post-anthesis fertilizer, Environment 3) 37oc/17oc day/night, well-watered, with post-anthesis fertilizer, Environment 4) 37oc/17oc day/night, well-watered, without post-anthesis fertilizer, Environment 5) 37oc/17oc day/night plus drought, with post-anthesis fertilizer, Environment 6) 37oc/17oc day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S.
	20, 24, 28, 30 DPA at USDA-ARS, Albany, CA. A cdna library was made using poly (A) RNA, and the cdna clones in the T1 Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."
BASE COUNT	95 a 180 c 197 g 66 t
ORIGIN	
Query Match	11 5%; Score 36.8; DB 14; Length 538;
Best Local Similarity	47.8%; Pred. No. 13; Mismatches 117; Indels 0; Gaps 0
Matches 107; Conservative	0; Mismatches 117; Indels 0; Gaps 0
11 GCGAGAGCGCGAGCCTGTGCGACACACAGCTCAGCGCGAGAGATGCGCG 70	
202 GCCAGGCGAGCGCGCGCGCGCGACAGCAAGCAGCGCGCGCGCGCGCTCT 261	
71 ACTGGCGCGCAGTGAACCGTCTGTCTGAACTCGTGTGCGCGCTCGCATGCGAAGG 130	

Db	262	CCGCCGTGCGCCCGCGCGGTGGGACGGGGCGGGGCGTAAACCCCTCTCCCGGGAGGG	321
QY	131	GGAAATGTTTCGTGCCCCGTGTGATGCGCTCACTACAGGAGATCTTCAGGACACCCACTGTGGAG	190
Db	322	ACGCCCTTCCGCGGGCTGATGATCATCAGCGCGCGCTTCTCTGTCGGCTCTCGGCTCTCGGCCCA	381
QY	191	AGGCCACAGACAGCTTGCGCGCGCAACACACTCAAGGACAGCTGGCGAG	234
Db	382	AGCCCTTCGCGCCCGCGGAGGACCACTCATCTACAGGGGCGGAG	425
RESULT 11			
LOCUS	AV651765	640 bp	mRNA linear EST 15-JAN-2002
DEFINITION	AV651765	GLC Homo sapiens cDNA clone G1CCTC09 3', mRNA sequence.	
ACCESSION	AV651765		
VERSION	AV651765.1	GI:9872779	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 640)		
JOURNAL	Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,		
MEDLINE	Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,		
COMMENT	Shen, K., Lu, G., Fu, G., Zhong, W., Xu, S., Gu, W., Huang, W., Zhao, X.,		
	Hu, G., Gu, J., Chen, Z., and Han, Z.		
	Insight into hepatocellular carcinoma genesis at transcriptome level		
	by comparing gene expression profiles of hepatocellular carcinoma		
	with those of corresponding noncancerous liver		
	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)		
	21625106		
	Contact: Zeguang Han		
	Chinese National Human Genome Center at Shanghai		
	351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai		
	201203, P. R. China		
	Tel: 86-21-50801919(ex.45)		
	Fax: 86-21-50801922		
	Email: hanzg@hgc.sh.cn		
FEATURES	This clone is available at CHGC in Shanghai.		
source	Location/Qualifiers		
	1..640		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="G1CCTC09"		
	/clone_1ib="G1C"		
	/tissue_type="corresponding non cancerous liver tissue"		
	/dev_stage="Adult"		
	/lab_host="SOLR"		
	/note="vector: pBluescript sk(-); Site_1: EcoRI; Site_2:		
	XhoI"		
BASE COUNT	139 a 203 c 196 g 100 t	2 others	
ORIGIN			
Query Match	11.5%;	Score 36.8;	DB 10; Length 640;
Best Local Similarity	48.5%;	Pred. No. 14;	
Matches 132; Conservative	0;	Mismatches 134;	Indels 6; Gaps 1
QY	8	ACTGCGAGGAGCCACGACTGCGCGGAACACAGCTCAGAGACGTGCGCGGAGAGATG	67
Db	374	ACCGCGAAGATGGAGGCGGTGCGCGCAAACTCCAMAGGGCGCGCGCAGAACTGC	433
QY	68	CCGACTTGGCGGCATGGAACCGGTGCTGTGAACTGCTGTGCGCGCTGCCTGACGAA	127
Db	434	CCGAGCTGCGAGGAGAGCTGAGCCCACTGGGCGAGAGATGNCNCGCCCGCGGCCC	493
QY	128	AGGGGAATGTTTCGTCGCTGATAGCTCACTACAGGAGATCTTCAGGACACCACTGCG	187
Db	494	ATTGTGACGCGCTGCGCAGCGCTCTTGCCCGCTTAAGGGACCACTTGGCCCAAGCTT	553
QY	188	AGGAGCCACGACGCTGCGCGAACAACAGCTCAGAGACGTGCGCGAGAGATGGCGACT	247

Db 554 TGGCCCGCCCGCTTGAAGGCTCTCAAGAGAAACCG- - - - - GGGCCAGATGGCCGACT 607

QY 248 TGGCCGCGATGGAACCGCTGCTGCTGAATC 279

Db 608 TCCACCGCAAGACCGACGCTTTTGACCGC 639

RESULT 12

LOCUS AM177874 304 bp mRNA linear EST 16-NOV-1999

DEFINITION IL3-HT0059-270899-010-H08 HT0059 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM177874

VERSION AM177874.1 GI:6443911

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 304)

TITLE HCGP <http://www.ludwig.org.br/ORESTES>.

JOURNAL The FAPESP/LICR Human Cancer Genome Project

COMMENT Unpublished (1999)

CONTACT: Simpson A.J.G.

LABORATORY of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL3&t2=IL3-HT0059-270899-010-H08&t3=1999-08-27&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 22

High quality sequence stop: 304.

Location/Qualifiers

1. 304

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0059"

/dev\_stage="Adult"

/note="Organ: head,neck; Vector: puc18; site.1: SmaI; site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (0.5. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 64 a 85 c 105 g 50 t

ORIGIN

Query Match 11.4%; Score 36.6; DB 10; Length 304;

Best Local Similarity 48.8%; Pred. No. 12;

Matches 99; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 73 TTGGCGGCGATGGAACCGTCTGCTGAATCTGTGCGCTGCGCATGACGAAAGGG 132

Db 74 TCGGCGCTCGATGACCTGTGGGGTGTGAAGTAGACTTCGACTCCACGAGAAAGGAA 133

QY 133 AATGTTTCTGCGCCGTTGATCGGCTCACTACAGGATCTCTAGGACCCACTGGCAGAG 192

Db 134 CAGCTCGCGAGCCGCGGTGGCAAGTGGTGACAAACCTATGACTTGTGACGCGCGGCGC 193

QY 193 GCGAGAGCTGCGCCGCAACAGCTTACAGAGCTGCGCGAAGAGTGGCCGACTTGGCG 252

Db 194 GGCATGTCAGAGCGGCGGCTACTCTCCAGAGAGCTTGGGCGAGATGGCTGAGCTGGGT 253

QY 253 GCGATGGAACCGTCTGTCTGA 275

Db 254 CTCATGGCGCTGGCGCTGCCAGA 276

RESULT 13

LOCUS AV627687 352 bp mRNA linear EST 15-DEC-2000

DEFINITION AV627687 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone ICL030e10\_r 5', mRNA sequence.

ACCESSION AV627687

VERSION AV627687.1 GI:10790321

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

AUTHORS 1 (bases 1 to 352)

TITLE Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y., and Tabata,S.

JOURNAL Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii

COMMENT DNA Res. 7 (5), 305-307 (2000)

CONTACT: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: [asamizu@kazusa.or.jp](mailto:asamizu@kazusa.or.jp), URL: <http://www.kazusa.or.jp/en/plant/>.

Location/Qualifiers

1. 352

/organism="Chlamydomonas reinhardtii"

/strain="C9"

/db\_xref="taxon:3055"

/clone="ICL030e10\_r"

/note="Vector: pBluescriptII SK-; Site.1: EcoRI; Site.2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

BASE COUNT 67 a 127 c 113 g 45 t

ORIGIN

Query Match 11.4%; Score 36.6; DB 10; Length 352;

Best Local Similarity 47.6%; Pred. No. 12;

Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 39 CAAGCTCAAGAGCTGGCGGAGAAAGGCGGACTGGCGCGCATGGAACCGTCTGTC 98

Db 1 CAACTCAAGCTGACACGCGGTGTGAGGCGGAGCTGGCTCATCGGCGAGAACTGCT 60

QY 99 TGAATCGTGTGGCGCTGCCATGACGAAAGGGAATGTTCTCGCCGTTGATCGCGTC 158

Db 61 CGACATGTTGCGCGCTCCGATGGCACACAGCAGCTCCAGCATCAAGCGGAGACGCT 120

QY 159 ACTACAGGATCTCTAGGACCCACTGCGAGAGGCCAGCAGCCTGCGCGAACAACGCT 218

Db 121 CTCACCGGCGAGTCTCCGCTCCGAGACGCGGCGCTGTGCTCCAGGCGCGTAGCAGTGT 180

QY 219 CAAGGAGCTGCGGAGAAAGTGGCGGCACTTGGCGCGCATGGAACCG 265

Db 181 CCGGGGTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 227

RESULT 14

LOCUS BG857743/c 707 bp mRNA linear EST 29-MAY-2001

DEFINITION 1024053604.xl C. reinhardtii CC-1690, normalized, lambda zap II

ACCESSION BG857743

VERSION BG857743.1 GI:14238927

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.



DB 1159 CCCACTGCGAGAGGCCA 1142

Search completed: June 2, 2003, 08:39:12  
Job time : 203 secsRESULT 15  
AAH14532/C

ID AAH14532 standard; CDNA; 4147 BP.

XX AC AAH14532;

XX DT 26-JUN-2001 (first entry)

XX DE Human CDNA sequence SEQ ID NO:12080.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PE 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PS Claim 8; SEQ ID 12080; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX SQ Sequence 4147 BP; 1078 A; 1016 C; 1118 G; 935 T; 0 other;

Query Match 5.6%; Score 18; DB 22; Length 4147;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 179 CCCACTGCGAGAGGCCA 196

|||||

polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: the sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1536 BP; 289 A; 500 C; 461 G; 286 T; 0 other;

Query Match 5.6%; Score 18; DB 23; Length 1536;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

18 GGCCAGCAGCCTGGCCGA 35  
 1254 GGCCAGCAGCCTGGCCGA 1237

# RESULT 13

ABO90147/c standard; DNA; 2595 BP.

ABO90147;

01-OCT-2002 (first entry)

M. capsulatus gene #132 for DNA array.

Micro array; gene; ds; differential expression; gene expression.

Methylococcus capsulatus.

WO200255655-A2.

18-JUL-2002.

14-JAN-2002; 2002WO-NO00019.

12-JAN-2001; 2001NO-0000235.

12-JAN-2001; 2001NO-0000239.

(UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.

(TIGR-) TIGR.

Bitkeland NK, Eldhammer I, Jonassen I, Jensen HB, Lien T;  
 PI Lillehaug JR, Lossius I, Eissen JA, Fraser CM, Durkin AS;  
 PI Salzberg SL;

WPI: 2002-557818/59.

Novel DNA array useful for determining differential expression of

Methylococcus capsulatus genes, comprises polynucleotides or

oligonucleotides representative for a selective number of Methylococcus

capsulatus genes

Claim 19; Page 115-116; 678pp; English.

The invention relates to a novel DNA array giving a representation of a

number of Methylococcus capsulatus genes. The method of the invention is

useful for determination of the differential expression of the genes of

M. capsulatus, and for studying gene expression on a genomic scale and in  
 CC gene expression assays of M. capsulatus genes. The sequences shown in  
 CC ABO90147-ABO91855 represent M. capsulatus genes for use in arrays of the  
 CC invention.

Sequence 2595 BP; 442 A; 737 C; 885 G; 531 T; 0 other;

Query Match 5.6%; Score 18; DB 24; Length 2595;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

37 CACAAGCTCAAGCAGCTG 54  
 429 CACAAGCTCAAGCAGCTG 412

RESULT 14  
 ABL23967 standard; DNA; 3873 BP.

ABL23967;

26-MAR-2002 (first entry)

Drosophila melanogaster genomic polynucleotide SEQ ID NO 23374.

Drosophila: developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ds.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI: 2001-656860/75.

New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signalling and cell-cell

interactions

Claim 1; SEQ ID NO 23374; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is

useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of

insecticides, therapeutics and pharmaceutical drugs. The invention

discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

sequences (ABL01840-ABL16175) and the encoded proteins

(ABBS7737-ABBS72072).

The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 3873 BP; 998 A; 1083 C; 996 G; 796 T; 0 other;

Query Match 5.6%; Score 18; DB 23; Length 3873;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

28 CTGGCCGAACACAGCTC 45  
 2956 CTGGCCGAACACAGCTC 2973



XX 12-OCT-2001; 2001WO-US31819.  
PF  
XX 12-OCT-2000; 2000US-240465P.  
PR  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
PA  
PI Summers AO, Cagulat JJ;  
XX  
XX WPI: 2002-435437/46.  
DR  
XX Novel non-naturally occurring recombinant DNA molecule encoding a  
PT chelon protein useful for binding divalent cation mercury from  
PR contaminated soil, water, aqueous medium including biological fluids -  
PF  
PS Disclosure; Page 24; 42pp; English.  
XX  
XX The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed.  
CC The nucleic acid encoding the chelon protein is useful for binding  
CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
CC metal ions from contaminated soil, ground water, hydroponic solutions or  
CC irrigation water of waste streams. The DNA of the invention, when  
CC immobilised onto a solid support, is useful for concentrating heavy metal  
CC ions from contaminated environment waste streams or contaminated  
CC aqueous medium including biological fluids. The nucleic acid, when  
CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
CC nonpathogenic), is suitable for use in the in vivo sequestration and  
CC elimination of mercuric ions from gastrointestinal tracts of animals or  
CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
CC molecules of the invention are also useful in water treatment resins.  
CC The nucleic acid of the invention is highly specific and binds divalent  
CC cation such as mercury or cadmium with high affinity. The present nucleic  
CC acid sequence represents synthetic product 1 forward PCR primer that was  
CC used in the methods of the invention for construction of pASK-MBD vector.  
XX  
XX Sequence 33 BP; 8 A; 8 C; 12 G; 5 T; 0 other;  
SQ  
Query Match 5.9%; Score 19; DB 24; Length 33;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACACACTGCGAGAG 19  
DB 15 ATGACACACTGCGAGAG 33  
11  
AAK87475  
ID AAK87475 standard; DNA; 1231 BP.  
XX  
XX AAK87475;  
AC  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42287.  
DE  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytosolic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157182-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 17-JAN-2001; 2001WO-US01354.  
PF

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205513.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227009.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.



XX Mercuric ion; contaminated soil; ground water; hydroponic solution;  
 KW Irrigation water; waste stream; contaminated aqueous medium;  
 KW biological fluid; gastrointestinal tract; chelon protein;  
 KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
 KW heavy metal binding protein; pASK-MBD; PCR; primer; ss.  
 OS Synthetic.

XX WO200230962-A2.

XX 18-APR-2002.

XX 12-OCT-2001; 2001WO-US31819.

XX 12-OCT-2000; 2000US-240465P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Summers AO, Caquiat JT;

XX WPI; 2002-435437/46.

XX Novel non-naturally occurring recombinant DNA molecule encoding a  
 PT chelon protein useful for binding divalent cation mercury from  
 PT contaminated soil, water, aqueous medium including biological fluids -  
 PS Disclosure; Page 24; 42pp; English.

XX The present invention relates to a new non-naturally occurring  
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
 CC which binds mercuric ions. The invention is useful for recombinantly  
 CC producing a protein in a host-cell, by infecting or transforming a host  
 CC cell capable of expressing a chelon coding sequence with a vector  
 CC comprising a promoter active in the host cell operably linked to a coding  
 CC region for the protein to produce a recombinant host cell and culturing  
 CC the recombinant host cell under conditions, where DNA is expressed.  
 CC The nucleic acid encoding the chelon protein is useful for binding  
 CC divalent mercuric ions, to take up, sequester and concentrate the heavy  
 CC metal ions from contaminated soil, ground water, hydroponic solutions or  
 CC irrigation water of waste streams. The DNA of the invention, when  
 CC immobilised onto a solid support, is useful for concentrating heavy metal  
 CC ions from contaminated environment waste streams or contaminated  
 CC aqueous medium including biological fluids. The nucleic acid, when  
 CC recombinantly expressed in enteric bacteria (which are nontoxicogenic and  
 CC nonpathogenic), is suitable for use in the in vivo sequestration and  
 CC elimination of mercuric ion from gastrointestinal tracts of animals or  
 CC humans exposed to toxic metal ions such as mercury and/or cadmium. The  
 CC molecules of the invention are also useful in water treatment resins.  
 CC The nucleic acid of the invention is highly specific and binds divalent  
 CC cation such as mercury or cadmium with high affinity. The present nucleic  
 CC acid sequence represents synthetic product 2 forward PCR primer that was  
 CC used in the methods of the invention for construction of pASK-MBD vector.

XX Sequence 30 BP; 7 A; 11 C; 8 G; 4 T; 0 other;

XX Query Match 9.3%; Score 30; DB 24; Length 30;  
 XX Best Local Similarity 100.0%; Pred. No. 7.8e-06;

XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX DB 160 CTACAGGATCTCTCGGACCCACTGCGAG 189  
 XX ||||||||||||||||||||||||||||  
 XX 1 CTACAGGATCTCTCGGACCCACTGCGAG 30

XX RESULT 9

XX ID AA058554 standard; DNA; 568 BP.

XX AC AA058554;

XX 16-SEP-1994 (first entry)

DE Mercury resistant control gene merR(1).  
 XX Resistance; mercury; selectable marker; Thiobacillus ferrooxidans;  
 KW transformation; detection; ss.  
 XX Thiobacillus ferrooxidans.

XX Key Location/Qualifiers  
 XX FT -35\_signal 23..28  
 XX FT -10\_signal 46..51  
 XX FT RBS 76..79  
 XX FT CDS 89..568  
 XX FT /tag= d  
 XX /product= merR(1) gene product.

XX JF06000083-A.

XX 11-JAN-1994.

XX 17-JAN-1991; 91JP-0018338.

XX 17-JAN-1991; 91JP-0018338.

XX (AKIT-) AKITA KEN.

XX (DOMA) DOMA MINING CO LTD.

XX WPI; 1994-077131/10.

XX P-PSDB; AAR49668.

XX Mercury resistant control gene merR and shuttle vector - for  
 PT enhanced expression of mercury resistance marker in transformed  
 PT Thiobacillus sp.

XX Claim 1; Page 2; 26pp; Japanese.

XX The mercury resistance genes can be used as selectable markers when  
 CC used to transform other bacteria.

XX Sequence 568 BP; 130 A; 150 C; 164 G; 124 T; 0 other;

XX Query Match 8.1%; Score 26; DB 15; Length 568;  
 XX Best Local Similarity 100.0%; Pred. No. 0.0009;

XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX DB 61 AAGATGGCGGACTTGGCGCGCATGA 86  
 XX ||||||||||||||||||||||||  
 XX 383 AAGATGGCGGACTTGGCGCGCATGA 408

XX RESULT 10

XX ID ABR52208 standard; DNA; 33 BP.

XX AC ABR52208;

XX 13-AUG-2002 (first entry)

XX Synthetic product 1 forward PCR primer for construction of pASK-MBD.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;  
 KW Irrigation water; waste stream; contaminated aqueous medium;  
 KW biological fluid; gastrointestinal tract; chelon protein;  
 KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
 KW heavy metal binding protein; pASK-MBD; PCR; primer; ss.

XX Synthetic.

XX WO200230962-A2.

XX 18-APR-2002.





	CC	The nucleic acid encoding the cheion protein is useful for purifying
	CC	divalent mercuric ions, to take up, sequester and concentrate the heavy
	CC	metal ions from contaminated soil, ground water, hydroponic solutions or
	CC	irrigation water of waste streams. The DNA of the invention, when
	CC	immobilised onto a solid support, is useful for concentrating heavy metal
	CC	ions from contaminated environment waste streams or contaminated
	CC	aqueous medium including biological fluids. The nucleic acid, when
	CC	recombinantly expressed in enteric bacteria (which are nontoxicogenic and
	CC	nonpathogenic), is suitable for use in the in vivo sequestration and
	CC	elimination of mercuric ion from gastrointestinal tracts of animals or
	CC	humans exposed to toxic metal ions such as mercury and/or cadmium. The
	CC	molecules of the invention are also useful in water treatment resins.
	CC	The nucleic acid of the invention is highly specific and binds diverent
	CC	cation such as mercury or cadmium with high affinity. The present nucleic
	CC	acid represents the adjacent plasmid region that encodes the variant Merr
	CC	protein of the invention.
	XX	
SQ	Sequence 509 BP; 139 A; 128 C; 140 G; 102 T; 0 other;	
Query Match	100.0%; Score 321; DB 24; Length 509;	
Best Local Similarity	100.0%; Pred. No. 1e-157;	
Matches 321; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ATGACACACTCGCAGAGGCCACGACCTGGCCGGAACACACACTCAAGAGCTGGCGAG 60	
Dd	138 ATGACACACTCGCAGAGGCCACGACCTGGCCGGAACACACACTCAAGAGCTGGCGAG 197	
OY	61 AAGATGGCGGACTTGGCGGCATGTGAACCCTGCTGTAACCTGTCGCTGCCCTGCCAT 120	
Dd	198 AAGATGGCGGACTTGGCGGCATGTGAACCCTGCTGTAACCTGTCGCTGCCCTGCCAT 257	
OY	121 GCACGAAGAAGGGGATGTTCTCTGCCCCGTGATCGCTACTACAGGATCCTCAGGCAC 180	
Dd	258 GCACGAAGAAGGGGATGTTCTCTGCCCCGTGATCGCTACTACAGGATCCTCAGGCAC 317	
OY	181 CACTGCGAGAGGCCACGACCTGGCGGAAACAACCTCAAGAGCTGGCGGAGAAGATG 240	
Dd	318 CACTGCGAGAGGCCACGACCTGGCGGAAACAACCTCAAGAGCTGGCGGAGAAGATG 377	
OY	241 GCCGACTTGCGCCGCGCATGGAACCCGTGCTGAACTCGTGTGGCGCTGCGCATGCACGA 300	
Dd	378 GCCGACTTGCGCCGCGCATGGAACCCGTGCTGAACTCGTGTGGCGCTGCGCATGCACGA 437	
OY	301 AAGGGGAATGTTCTCTGCCCG 321	
Dd	438 AAGGGGAATGTTCTCTGCCCG 458	
RESULT 4		
ID	ABK52206 standard; DNA; 435 BP.	
XX	ABK52206;	
XX	13-AUG-2002 (first entry)	
DE	DNA encoding Shigella flexneri wild-type Merr protein.	
XX		
KW	Mercuric ion; contaminated soil; ground water; hydroponic solution;	
KW	irrigation water; waste stream; contaminated aqueous medium;	
KW	biological fluid; gastrointestinal tract; cheion protein;	
KW	enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;	
KW	heavy metal binding protein; Merr; gene; ds.	
OS	Shigella flexneri.	
XX		
FH	Key Location/Qualifiers	
FT	CDS 1..435	
FT	/tag= a	
FT	/product= "Shigella flexneri wild-type Merr protein"	
PN	MO200230962-A2.	

XX 12-OCT-2000; 2000US-240465P.  
PR (URGE-) UNIV GEORGIA RES FOUND INC  
PA  
XX  
XX Summers AO, Caguiat JJ;  
PI  
XX  
XX  
DR MPI: 2002-435437/46.  
DR P-PSDB; AA097552.  
XX

Novel non-naturally occurring recombinant DNA molecule encoding a chelator protein useful for binding divalent cation mercury from contaminated soil, water, aqueous medium including biological fluids -  
Disclosure: Page 21; 42pp. English.

The present invention relates to a new non-naturally occurring recombinant DNA molecule comprising a sequence encoding a chelon protein which binds mercuric ions. The invention is useful for recombinantly producing a protein in a host-cell, by infecting or transforming a host cell capable of expressing a chelon coding sequence with a vector comprising a promoter active in the host cell operably linked to a coding region for the protein to produce a recombinant host cell and culturing the recombinant host cell under conditions, where DNA is expressed. The nucleic acid encoding the chelon protein is useful for binding divalent mercuric ions, to take up, sequester and concentrate the heavy metal ions from contaminated soil, ground water, hydroponic solutions or irrigation water of waste streams. The DNA of the invention, when immobilised onto a solid support, is useful for concentrating heavy metal ions from contaminated environment waste streams or contaminated aqueous medium including biological fluids. The nucleic acid, when recombinantly expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic), is suitable for use in the in vivo sequestration and elimination of mercuric ion from gastrointestinal tracts of animals or humans exposed to toxic metal ions such as mercury and/or cadmium. The molecules of the invention are also useful in water treatment resins. The nucleic acid of the invention is highly specific and binds divalent cation such as mercury or cadmium with high affinity. The present nucleic acid sequence encodes the synthetic Merr chelon variant protein of the invention. This sequence encodes one of the heavy metal binding proteins termed chelons of the invention.

Sequence 321 BP; 72 A; 94 C; 103 G; 52 T; 0 other;

Query Match	100.0%;	Score 321;	DB 24;	Length 321;
Best Local Similarity	100.0%;	Pred No 1;	1st 157	

Matches	321	Conservative	0	Mismatches	0	Indels	0	Gaps	0
---------	-----	--------------	---	------------	---	--------	---	------	---

1 ATGACACACTGGCAGGAGGCCACGACCTGGCCGAACAAGCTCAGGACGTCCGAG 60

1 ATGACACACTGCGAGGAGGCCAGCAGCCTGGCCGAACACAGCTCAAGGAGCTGGCGAG 60

61 ||||| 12

[illegible][illegible]

181 CACTGAGAGAGGCCACCGCCTCCCCCAAAATTCCTAATTTT  
-----  
180 TCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT


181 CACTCGAGGAGGCACCTAGCCTTGGCCCAACACACCTAACATTCCTCCCCCATTT  
24

241 GCCGACTTGGCGCCGCATGGAACCGGTGCTGTCTGAACTCGTCTGCGCGCTTCAATCAAGAG, 300

241 GCGACTTGGCGCCGATGGAACCGTGCTCTTGAACTGTTGCGGCTGCAATGCAACA 300

301 AAGGGAATGTTTCCTGCCCG 321

301 AAGGGAATGTTCTCCTGCCCG 321

RESULT 2  
ABK52212  
TD ABK52212

AC ABK52212;

DT 13-AUG-2002 (first entry)  
 YY

DE DNA encoding PASK-MBD gene.  
XX

mercuric ion; contaminated soil; ground water; hydroponic solution; irrigation water; waste stream; contaminated

enteric bacteria; toxic metal ion; mercury; cadmium; aluminum

XX *moda binding protein*; PCR; pASK-MBD; gene; ds.

XX

[illegible]

XX	13-0000-2001	2001-0000-2001
DE	13-0000-2001	2001-0000-2001

XX	12-OCT-2000
PR	2000MS-2404552

XX PA (HYGE-) TINTV GEORCTA RES EOTUD TNG

Summers AO, Gaudiat JT.  
PI

WPI; 2002-435437/46

PT Novel non-naturally occurring recombinant DNA molecule encoding a

contaminated soil, water, aqueous medium including biological fluids -

PS Disclosure; Page 24; 42pp; English.

The present invention relates to a new non-naturally occurring recombinant DNA molecule comprising a sequence encoding a chelon protein which binds mercuric ions. The invention is useful for recombinantly producing a protein in a host-cell, by infecting or transforming a host cell capable of expressing a chelon coding sequence with a vector comprising a promoter active in the host cell operably linked to a coding region for the protein to produce a recombinant host cell and culturing the recombinant host cell under conditions, where DNA is expressed. The nucleic acid encoding the chelon protein is useful for binding divalent mercuric ions, to take up, sequester and concentrate the heavy metal ions from contaminated soil, ground water, hydroponic solutions or irrigation water of waste streams. The DNA of the invention, when immobilised onto a solid support, is useful for concentrating heavy metal ions from contaminated environmental waste streams or contaminated aqueous medium including biological fluids. The nucleic acid, when recombinantly expressed in enteric bacteria (which are non-toxicogenic and nonpathogenic), is suitable for use in the *in vivo* sequestration and elimination of mercuric ion from gastrointestinal tracts of animals or humans exposed to toxic metal ions such as mercury and/or cadmium. The molecules of the invention are also useful in water treatment resins. The nucleic acid of the invention is highly specific and binds divalent cation such as mercury or cadmium with high affinity. The present nucleic acid sequence represents the PASK-MBD gene, as described in the invention.

Sequence 354 BP; 83 A; 103 C; 111 G; 57 T; 0 other;

Query Match	100.08;	Score 321;	DB 24;	Length 354;
Best Local Similarity	100.08			

```

matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 06:54:09 ; Search time 199 Seconds  
(without alignments)  
3632.620 Million cell updates/sec

Title: US-09-977-137A-3

Perfect score: 321  
Sequence: 1 atgacacactggagagagc.....aggaggaattctctgccgcg 321

Scoring table: OLIGO\_MDC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 segs, 112599159 residues

Word size: 0  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: N\_Geneseq\_101002: \*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	100.0	321	24	ABK52207
2	321	100.0	354	24	ABK52212
3	321	100.0	509	24	ABK52213
4	161	50.2	435	24	ABK52206
5	161	50.2	7076	13	AAO20186
6	161	50.2	7076	16	AAO86906
7	33	10.3	33	24	ABK52209
8	30	9.3	30	24	ABK52210
9	26	8.1	568	15	AAO58554

10	19	5.9	33	24	ABK52208
11	18	5.6	1231	22	AAK87475
12	18	5.6	1536	23	AAK82310
13	18	5.6	2595	24	ABO90147
14	18	5.6	3873	23	ABL23967
15	18	5.6	4147	22	AAH14532
16	18	5.6	4802	23	ABV24968
17	18	5.6	4802	23	ABV25047
18	18	5.6	4802	23	ABV25170
19	18	5.6	4802	23	ABL23966
20	17	5.3	1356	22	AAH51971
21	17	5.3	2043	24	AAO90098
22	17	5.3	2505	22	AAH48054
23	17	5.3	4911	23	ABL10841
24	17	5.3	7886	23	ABL10840
25	17	5.3	4403765	22	AAI99683
26	17	5.3	4411529	22	AAI99682
27	16	5.0	261	19	AAV38052
28	16	5.0	262	19	AAV38049
29	16	5.0	266	19	AAV38055
30	16	5.0	278	19	AAV38048
31	16	5.0	289	19	AAV38050
32	16	5.0	299	21	AAV38051
33	16	5.0	335	19	AAV38051
34	16	5.0	384	22	AAV65140
35	16	5.0	397	22	AAV65139
36	16	5.0	422	22	AAV65139
37	16	5.0	438	22	AAH48287
38	16	5.0	476	11	AAO04716
39	16	5.0	477	10	AAH91857
40	16	5.0	477	13	AAQ29741
41	16	5.0	508	24	ABN73421
42	16	5.0	526	21	AAI10058
43	16	5.0	570	21	AAI14731
44	16	5.0	815	22	AAH01852
45	16	5.0	866	22	AAH04301

#### ALIGNMENTS

RESULT 1  
ABK52207  
ID ABK52207 standard; DNA: 321 BP.

AC ABK52207:  
XX  
DT 13-AUG-2002 (first entry)

DE DNA encoding synthetic Merr chelon variant.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;  
XX Irritation water; waste stream; contaminated aqueous medium;  
KW biological fluid; gastrointestinal tract; chelon protein;  
KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;  
KW heavy metal binding protein; Merr; gene; ds.

XX Synthetic.

OS  
XX  
FH Key  
FT 1..321  
FT CDS  
Location/Qualifiers

FT /partial  
FT /product= "Synthetic Merr chelon variant"  
FT /note= "This sequence encodes the first 107 amino acids  
of the synthetic Merr chelon variant. This  
sequence lacks a stop codon"

XX WO200230962-A2.

XX 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31819.

Synthetic product  
Human immune/haema  
DNA encoding novel  
M. capsulatus gene  
Drosophila melanog  
Human cDNA sequenc  
Human prostate exp  
Human prostate exp  
Drosophila melanog  
Mycobacterium tube  
M. capsulatus gene  
Murine protein cod  
Drosophila melanog  
Mycobacterium tube  
Mycobacterium tube  
Human stem cell an  
Human stem cell an  
Human stem cell an  
Human stem cell an  
Human stem cell an  
Human secreted pro  
Novel human polyu  
Novel human polyu  
Human polynucleoti  
Human colon cancer  
Nucleotide sequenc  
Synthetic human fo  
Synthetic basic fg  
Bovine embryonic g  
Human secreted pro  
Aspergillus oryzae  
Arcanobacterium ha  
Human cDNA clone (

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 06:52:59 ; Search time 1797 Seconds

(without alignments)  
5198.665 Million cell updates/sec

Title: US-09-977-137A-3

Perfect score: 321

Sequence: 1 atgacacactgcagagagc.....agggaatgttctcgcgcg 321

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_ov:\*  
22: em\_or:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pln:\*  
35: em\_hlg\_rod:\*  
36: em\_hlg\_man:\*  
37: em\_hlg\_vtc:\*  
38: em\_sy:\*  
39: em\_hlg\_hum:\*  
40: em\_hlg\_mus:\*  
41: em\_hlg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	321	100.0	321	6	AX473152	AX473152 Sequence
2	321	100.0	354	6	AX473166	AX473166 Sequence
3	321	100.0	509	6	AX473167	AX473167 Sequence
4	161	50.2	435	6	AX473150	AX473150 Sequence
5	161	50.2	445	1	PFL422224	AJ422224 Pseudomon
6	161	50.2	445	1	PFL422225	AJ422225 Pseudomon
7	161	50.2	3763	1	ARMERGE4	233486 A. faecalis
8	161	50.2	5747	1	RI004	K03089 Plasmid NR1
9	161	50.2	19672	1	AF071413	AF071413 Assembled
10	161	50.2	26652	1	AF188331	AF188331 Shigella
11	161	50.2	94281	1	AP000342	AP000342 Plasmid R
12	161	50.2	218160	1	STYPPHGM1	AL513383 Salmonella
13	161	50.2	1696	1	AFMERG5	233485 K. oxytoca (
14	161	50.2	1696	1	EC077087	U77087 Escherichia
15	161	50.2	2301	1	AB013925	AB013925 Pseudomon
16	161	50.2	12373	1	D83080	D83080 Pseudomonas
17	161	50.2	5505	1	PSPI7897	Y17897 Pseudomonas
18	161	50.2	12373	1	PPS304453	AJ304453 Sequence
19	161	50.2	55578	33	AX473163	AF092069 Pseudomon
20	161	50.2	333	1	AF092069	233488 E. aerogenes
21	161	50.2	516	1	FAEMERGE8	AX473164 Sequence
22	161	50.2	30	6	AX473164	233490 P. fluoresce
23	161	50.2	516	1	FAEMERGE10	233489 P. sp. (SB3M
24	161	50.2	516	1	FAEMERGE9	233482 A. calcoacet
25	161	50.2	516	1	FAEMERGE2	233483 A. calcoacet
26	161	50.2	516	1	FAEMERGE3	233486 E. coli (E66157
27	161	50.2	516	1	FAEMERGE7	233487 E. coli (E66157
28	161	50.2	516	1	FAEMERGE6	233488 E. coli (E66157
29	161	50.2	516	1	FAEMERGE5	233489 E. coli (E66157
30	161	50.2	516	1	FAEMERGE4	233490 E. coli (E66157
31	161	50.2	516	1	FAEMERGE3	233491 E. coli (E66157
32	161	50.2	516	1	FAEMERGE2	233492 E. coli (E66157
33	161	50.2	516	1	FAEMERGE1	233493 E. coli (E66157
34	161	50.2	516	1	FAEMERGE0	233494 E. coli (E66157
35	161	50.2	516	1	FAEMERGE0	233495 E. coli (E66157
36	161	50.2	516	1	FAEMERGE0	233496 E. coli (E66157
37	161	50.2	516	1	FAEMERGE0	233497 E. coli (E66157
38	161	50.2	516	1	FAEMERGE0	233498 E. coli (E66157
39	161	50.2	516	1	FAEMERGE0	233499 E. coli (E66157
40	161	50.2	516	1	FAEMERGE0	233500 E. coli (E66157
41	161	50.2	516	1	FAEMERGE0	233501 E. coli (E66157
42	161	50.2	516	1	FAEMERGE0	233502 E. coli (E66157
43	161	50.2	516	1	FAEMERGE0	233503 E. coli (E66157
44	161	50.2	516	1	FAEMERGE0	233504 E. coli (E66157
45	161	50.2	516	1	FAEMERGE0	233505 E. coli (E66157

## ALIGNMENTS

RESULT 1	AX473152	321 bp	DNA	linear	PAT 09-AUG-2002
LOCUS	AX473152				
DEFINITION	Sequence 3 from Patent WO0230962.				
ACCESSION	AX473152				
VERSION	AX473152.1				
KEYWORDS	GI:22207871				
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Summers, A.O. and Caguiat, J.J.				
TITLE	Metal binding proteins, recombinant host cells and methods				
JOURNAL	Patent: WO 0230962-A.3 18-APR-2002;				
	UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)				

FEATURES	Location/Qualifiers
source	1..321 /organism="synthetic construct" /db_xref="taxon:32630" /note="chelon"
BASE COUNT	72 a 94 c 103 g 52 t
ORIGIN	
Query Match	100.0%; Score 321; DB 6; Length 321;
Best Local Similarity	100.0%; Pred. No. 1.3e-152; Indels 0; Gaps 0;
Matches 321; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ATGACACACTGCGAGAGGCGCCAGACGCTGGCCGACACAAAGCTCAAGAGCTGGCGGAG 60
Db	1 ATGACACACTGCGAGAGGCGCCAGACGCTGGCCGACACAAAGCTCAAGAGCTGGCGGAG 60
OY	61 AAGATGGCCGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCAT 120
Db	61 AAGATGGCCGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCAT 120
OY	121 GCACGAAAGGGGAATGTTTCCTGCGCCCTGTGATCCGTCCTACTACAGAGATCTCAGGCACC 180
Db	121 GCACGAAAGGGGAATGTTTCCTGCGCCCTGTGATCCGTCCTACTACAGAGATCTCAGGCACC 180
OY	181 CACTGCGAGAGGCGCCAGCAGCTGCGCCGACACAAAGCTCAAGAGAGTGGCGCGAAGATG 240
Db	181 CACTGCGAGAGGCGCCAGCAGCTGCGCCGACACAAAGCTCAAGAGAGTGGCGCGAAGATG 240
OY	241 GCGGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCATGCACGA 300
Db	241 GCGGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCATGCACGA 300
OY	301 AAGGGGAATGTTTCCTGCGCCG 321
Db	301 AAGGGGAATGTTTCCTGCGCCG 321
RESULT 2	
LOCUS	AX473166 354 bp DNA linear PAT 09-AUG-2002
DEFINITION	Sequence 17 from Patent WO0230962.
ACCESSION	AX473166
VERSION	AX473166.1 GI:22207876
KEYWORDS	
SOURCE	synthetic construct.
ORGANISM	synthetic construct.
REFERENCE	artificial sequences.
AUTHORS	1
TITLE	Summers, A.O. and Caguiat, J.J.
JOURNAL	Metal binding proteins, recombinant host cells and methods
	Patent: WO 0230962-A 17 18-Apr-2002;
	UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)
FEATURES	Location/Qualifiers
source	1..354 /organism="synthetic construct" /db_xref="taxon:32630" /note="sequence encoding chelon"
BASE COUNT	83 a 103 c 111 g 57 t
ORIGIN	
Query Match	100.0%; Score 321; DB 6; Length 354;
Best Local Similarity	100.0%; Pred. No. 1.3e-152;
Matches 321; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ATGACACACTGCGAGAGGCGCCAGACGCTGGCCGACACAAAGCTCAAGAGCTGGCGGAG 60
Db	1 ATGACACACTGCGAGAGGCGCCAGACGCTGGCCGACACAAAGCTCAAGAGCTGGCGGAG 60
OY	61 AAGATGGCCGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCAT 120
Db	61 AAGATGGCCGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCAT 120
OY	121 GCACGAAAGGGGAATGTTTCCTGCGCCCTGTGATCCGTCCTACTACAGAGATCTCAGGCACC 180
Db	121 GCACGAAAGGGGAATGTTTCCTGCGCCCTGTGATCCGTCCTACTACAGAGATCTCAGGCACC 180
OY	181 CACTGCGAGAGGCGCCAGCAGCTGCGCCGACACAAAGCTCAAGAGAGTGGCGCGAAGATG 240
Db	181 CACTGCGAGAGGCGCCAGCAGCTGCGCCGACACAAAGCTCAAGAGAGTGGCGCGAAGATG 240
OY	241 GCGGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCATGCACGA 300
Db	241 GCGGACTTGGCGCGCATGTGAACCCGTGCTGTGAACCTGTGGCCCTGCCATGCACGA 300
OY	301 AAGGGGAATGTTTCCTGCGCCG 321
Db	301 AAGGGGAATGTTTCCTGCGCCG 321

Db	121	GCACGAAAGGGGAATGTTCTCCTGCCCGTTGATCGCTACTACAGGATCCTCAGGCACC	180
QY	181	CACGTGCGAGGAGGCCAGCAGCCTGGCCGCAACACACAGCTCAAGACGTGCGGANAAGATC	240
Db	181	CACGTGCGAGGAGGCCAGCAGCCTGGCCGCAACACACAGCTCAAGACGTGCGGANAAGATG	240
QY	241	GCCGACTTGGGCGCGCATGGAAACCGCGATCGTGAACCTGTGTGCGCCTGCATGCACGA	300
Db	241	GCCGACTTGGGCGCGCATGGAAACCGCGATCGTGAACCTGTGTGCGCCTGCATGCACGA	300
QY	301	AAGGGGAATGTTCTCCTGCCCG	321
Db	301	AAGGGGAATGTTCTCCTGCCCG	321
RESULT 3			
LOCUS	AX473167	509 bp	DNA
DEFINITION	Sequence 18 from Patent WO0230962.		linear
ACCESSION	AX473167		
VERSION	AX473167.1		GI:22207877
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE	1		
AUTHORS	Summers,A.O. and Caguiat,J.J.		
TITLE	Metal binding proteins, recombinant host cells and methods		
JOURNAL	Patent: WO 0230962-A, 18 18-APR-2002;		
FEATURES	UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)		
source	Location/Qualifiers		
	1..509		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/note="sequence encoding chelon flanked by sequences		
	derived from plasmid"		
BASE COUNT	139 a	128 c	140 g 102 t
ORIGIN			
Query Match		100.0%;	Score 321; DB 6; Length 509;
Best local Similarity		100.0%;	Pred. No. 1.2e-152;
Matches	321; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
QY	1	ATGACACACTGCGAGGAGGCCAGCAGCCTGGCCGCAACACACAGCTCAAGGACGTGCGCAG	60
Db	138	ATACACACTGCGAGGAGGCCAGCAGCAGCCTGGCCGCAACACACAGCTCAAGGACGTGCGCAG	197
QY	61	AAGATGCGCAGCTTGGCGCGCATGGAAACCGTGTCTGAACCTCGTGTGCGCGTGCAT	120
Db	198	AAGATGCGCAGCTTGGCGCGCATGGAAACCGTGTCTGAACCTCGTGTGCGCGTGCAT	257
QY	121	GCACGAAGGGGATGTTCTCCTGCCCGTTGATCGCTACTACAGGATCCTCAGGCACC	180
Db	258	GCACGAAGGGGATGTTCTCCTGCCCGTTGATCGCTACTACAGGATCCTCAGGCACC	317
QY	181	CACGTGCGAGGAGGCCAGCAGCCTGGCCGCAACACACAGCTCAAGGACGTGCGGANAAGATG	240
Db	318	CACGTGCGAGGAGGCCAGCAGCCTGGCCGCAACACACAGCTCAAGGACGTGCGGANAAGATG	377
QY	241	GCCGACTTGGGCGCGCATGGAAACCGTGTCTGAACCTGTGTGCGCCTGCATGCACGA	300
Db	378	GCCGACTTGGGCGCGCATGGAAACCGTGTCTGAACCTGTGTGCGCCTGCATGCACGA	437
QY	301	AAGGGGAATGTTCTCCTGCCCG	321
Db	438	AAGGGGAATGTTCTCCTGCCCG	458
RESULT 4			
LOCUS	AX473150	435 bp	DNA
DEFINITION	Sequence 1 from Patent WO0230962.		linear
			PAT 09-AUG-2002



ACCESSION AX473150  
VERSION AX473150.1 GI:22207870  
KEYWORDS  
SOURCE  
ORGANISM *Shigella flexneri*.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
*Shigella*.

REFERENCE  
1 Summers, A.O. and Caguiat, J.J.  
Metal binding proteins, recombinant host cells and methods  
Patent: WO 0230962-A1 18-APR-2002;  
UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US)  
Location/Qualifiers  
1. .445  
/organism="Shigella flexneri"  
/db\_xref="taxon:623"  
/note="Tn21 of Plasmid R100"

BASE COUNT 95 a 113 c 146 g 81 t  
ORIGIN

Query Match 50.2%; Score 161; DB 6; Length 435;  
Best Local Similarity 100.0%; Pred. No. 3.9e-71;  
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CACTGCGAGGAGCCGACGCTGCGCGAAGCTCAAGAGAGCGCGAGAGATG 66  
DB 241 CACTGCGAGGAGCCGACGCTGCGCGAAGCTCAAGAGAGCGCGAGAGATG 300  
QY 67 GCCGACTTGGCGCGATGGAACCGTGTCTGACTGCTGTGCGCGCTGCATGCAG 126  
DB 301 GCCGACTTGGCGCGATGGAACCGTGTCTGACTGCTGTGCGCGCTGCATGCAG 360  
QY 127 AAGGGGAATGTTCTCTGCGCGCTGTGATCGGCTCACTACAGGG 167  
DB 361 AAGGGGAATGTTCTCTGCGCGCTGTGATCGGCTCACTACAGGG 401

RESULT 5  
PFL422224 445 bp DNA linear BCT 07-JUL-2002  
LOCUS  
DEFINITION *Pseudomonas fluorescens* (strain KHP22) transposon Tn5041g, partial  
sequence (region of insertion of Tn21 homologue).  
ACCESSION AJ422224  
VERSION AJ422224.1 GI:19848211  
KEYWORDS  
SOURCE *Pseudomonas fluorescens*.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; *Pseudomonadaceae*;  
*Pseudomonas*.

REFERENCE  
1 Kholodii, G.Y., Gorlenko, Z.M., Mindlin, S.Z. and Nikiforov, V.G.  
Distribution of distinct microvariants of Tn5041 in environmental  
bacteria  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 445)  
AUTHORS Kholodii, G.Y.  
TITLE Direct Substitution  
JOURNAL Submitted (18-DEC-2001) Kholodii G.Y., Russian Academy of Sciences,  
Institute of Molecular Genetics, Kurchatov Sq. 2, Moscow 123182,  
RUSSIA  
Related sequence Y09209.  
Location/Qualifiers  
1. .445  
/organism="Pseudomonas fluorescens"  
/strain="KHP22"  
/db\_xref="taxon:294"  
/sex="Tn21 (narrow range)"  
/plasmid="pKHP22"  
/country="Kyrgyzstan:Khalidarkan"  
/note="isolated from mercury mine, ore"  
repeat\_region 1. .445  
repeat\_unit 31. .35

/note="the sequence that has been duplicated in the course  
of insertion of the Tn21 homologue"  
/rpt\_type=DIRECT  
36. .>445  
repeat\_region /note="Tn21 homologue lacking the integrin"  
/transposon="Tn21deltaTn21"  
36. .73  
repeat\_unit /note="Tn21 subgroup transposons"  
/rpt\_type=INVERTED  
/rpt\_type=TERMINAL  
complement(69. .445)  
/gene="mer"  
/pseudo  
complement(69. .>445)  
/gene="mer"  
/function="activator/repressor"  
/pseudo  
/codon\_start=3  
/transl\_table=11  
/product="Mer protein"

BASE COUNT 88 a 149 c 119 g 89 t  
ORIGIN

Query Match 50.2%; Score 161; DB 1; Length 445;  
Best Local Similarity 100.0%; Pred. No. 3.9e-71;  
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CACTGCGAGGAGCCGACGCTGCGCGAAGCTCAAGAGAGCGCGAGAGATG 66  
DB 263 CACTGCGAGGAGCCGACGCTGCGCGAAGCTCAAGAGAGCGCGAGAGATG 204  
QY 67 GCCGACTTGGCGCGATGGAACCGTGTCTGACTGCTGTGCGCGCTGCATGCAG 126  
DB 203 GCCGACTTGGCGCGATGGAACCGTGTCTGACTGCTGTGCGCGCTGCATGCAG 144  
QY 127 AAGGGGAATGTTCTCTGCGCGCTGTGATCGGCTCACTACAGGG 167  
DB 143 AAGGGGAATGTTCTCTGCGCGCTGTGATCGGCTCACTACAGGG 103

RESULT 6  
PFL422225/c 445 bp DNA linear BCT 07-JUL-2002  
LOCUS  
DEFINITION *Pseudomonas fluorescens* (strain KHP25) transposon Tn5041g, partial  
sequence (region containing a Tn21 homologue).  
ACCESSION AJ422225  
VERSION AJ422225.1 GI:19848212  
KEYWORDS  
SOURCE *Pseudomonas fluorescens*.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; *Pseudomonadaceae*;  
*Pseudomonas*.

REFERENCE  
1 Kholodii, G.Y., Gorlenko, Z.M., Mindlin, S.Z. and Nikiforov, V.G.  
Distribution of distinct microvariants of Tn5041 in environmental  
bacteria  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 445)  
AUTHORS Kholodii, G.Y.  
TITLE Direct Substitution  
JOURNAL Submitted (18-DEC-2001) Kholodii G.Y., Russian Academy of Sciences,  
Institute of Molecular Genetics, Kurchatov Sq. 2, Moscow 123182,  
RUSSIA  
Location/Qualifiers  
1. .445  
/organism="Pseudomonas fluorescens"  
/strain="KHP25"  
/db\_xref="taxon:294"  
/sex="Tn21 (narrow range)"  
/plasmid="pKHP22"  
/country="Kyrgyzstan:Khalidarkan"  
/note="isolated from mercury mine, ore"  
repeat\_region 1. .445  
repeat\_unit 31. .35

```

repeat_unit
    /transposon="Tn5041G"
    31..35
    /note="the sequence that has been duplicated in the course
    of insertion of the tn21 homologue"
    /rpl_type=DIRECT
    36..>445
    /note="tn21 homologue taking the integron"
    /transposon="Tn21deltaIn2"
    36..73
    /note="tn21 subgroup transposons"
    /rpl_type=INVERTED
    /rpl_type=TERMINAL
    complement(69..445)
    /gene="merR"
    complement(69..>445)
    /gene="merR"
    /function="activator/repressor"
    /codon_start=3
    /transl_table=11
    /product="MerR protein"
    /protein_id="CAD19592.1"
    /db_xref="GI:19848213"
    /translation="VTIRYORKGLPEPDKPYGSIIRYGEADVVRVFKSAORLG
    ISLDEIAELRLDDCHCEPASLAEHKLKLDVREKMDLAMEYVLSLVCACHARKG
    NVSPILASLOGEGLARSAMP"
BASE COUNT      88 a      149 c      119 g      89 t
ORIGIN
Query Match      50.2%; Score 161; DB 1; Length 445;
Best Local Similarity 100.0%; Pred. No. 3.9e-71;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CACTGCGAGAGGCGCAGACCTGCGCGACACAAAGCTCAAGAGCGCGGAGAAAGATG 66
    |||||||
DB 263 CACTGCGAGAGGCGCAGACCTGCGCGACACAAAGCTCAAGAGCGCGGAGAAAGATG 204
    |||||||
QY 67 GCCGACTTGGCGCGCATGGAACCGTGTCTGTAACCTGTCGCGCCATGACGACA 126
    |||||||
DB 203 GCCGACTTGGCGCGCATGGAACCGTGTCTGTAACCTGTCGCGCCATGACGACA 144
    |||||||
QY 127 AAGGGGAATGTTCTGTCGCCGTTGATCGGTCACACAGG 167
    |||||||
DB 143 AAGGGGAATGTTCTGTCGCCGTTGATCGGTCACACAGG 103
    |||||||

RESULT 7
ARMERGE4      516 bp      DNA      linear      BCF 31-DEC-1994
LOCUS      A.faecalis (SE20MERR) merR gene for regulatory protein.
DEFINITION
ACCESSION      233484
VERSION      233484.1 GI:607038
KEYWORDS      merR; regulatory protein.
SOURCE      Alcaligenes faecalis.
ORGANISM      Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Alcaligenes.
REFERENCE      1 (bases 1 to 516)
AUTHORS      Osborn,A.M., Bruce,R.D., Strike,P. and Ritchie,D.A.
TITLE      Sequence Conservation between Regulatory Mercury Resistance Genes
            From Mercury Polluted and Pristine Environments
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 516)
AUTHORS      Osborn,A.M.
TITLE      Direct Submission
JOURNAL      Submitted (13-MAY-1994) Osborn A. M., University of Liverpool,
            Genetics and Microbiology, Liverpool, Merseyside, UK, L69 3BX
            Location/Qualifiers
FEATURES
    source
        1..516
        /organism="Alcaligenes faecalis"
        /isolate="SE20"
        /db_xref="taxon:511"
        /clone="pSE20R1, pSE20R2"
        complement(13..18)
-35_signal

```

```

-10_signal
    /citation=[1]
    complement(28..33)
-10_signal
    /citation=[1]
    36..41
-35_signal
    /citation=[1]
    61..66
gene
    /citation=[1]
    72..506
    /gene="merR"
    72..506
    /gene="merR"
CDS
    72..506
    /gene="merR"
    /citation=[1]
    /codon_start=1
    /transl_table=11
    /product="regulatory protein"
    /protein_id="CA483892.1"
    /db_xref="GI:607039"
    /db_xref="SPIREMBL:Q44190"
    /translation="MENNLENTLIGVFAKANGVETIRFYORKGLPEPDKPYGSI
    RYGEADVVRVFKSAORLGFSLDEIAELRLDDCHCEPASLAEHKLKLDVREKMD
    LAMEYVLSLVCACHARKNVSPILASLOGEGLARSAMP"
BASE COUNT      112 a      130 c      165 g      109 t
ORIGIN
Query Match      50.2%; Score 161; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 3.8e-71;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CACTGCGAGAGGCGCAGACCTGCGCGACACAAAGCTCAAGAGCGCGGAGAAAGATG 66
    |||||||
DB 312 CACTGCGAGAGGCGCAGACCTGCGCGACACAAAGCTCAAGAGCGCGGAGAAAGATG 371
    |||||||
QY 67 GCCGACTTGGCGCGCATGGAACCGTGTCTGTAACCTGTCGCGCCATGACGACA 126
    |||||||
DB 372 GCCGACTTGGCGCGCATGGAACCGTGTCTGTAACCTGTCGCGCCATGACGACA 431
    |||||||
QY 127 AAGGGGAATGTTCTGTCGCCGTTGATCGGTCACACAGG 167
    |||||||
DB 432 AAGGGGAATGTTCTGTCGCCGTTGATCGGTCACACAGG 472
    |||||||

RESULT 8
NR1MER/8
LOCUS      Plasmid NRI mercury resistance (mer) operon.
DEFINITION
ACCESSION      K03089
VERSION      K03089.1 GI:150389
KEYWORDS      merA gene; merR gene; merT gene; mercuric reductase.
SOURCE      Plasmid NRI (IncFII) from E.coli DNA, clone PDB7.
ORGANISM      plasmids.
REFERENCE      1 (bases 1 to 3763)
AUTHORS      Barrineau,P., Gilbert,P., Jackson,W.J., Jones,C.S., Summers,A.O.
            and Wisdom,S.
TITLE      The DNA sequence of the mercury resistance operon of the IncFII
            plasmid NRI
JOURNAL      J. Mol. Appl. Genet. 2 (6), 601-619 (1984)
MEDLINE      85159407
PUBMED      6530603
REFERENCE      2 (bases 1 to 3763)
AUTHORS      Summers,A.O.
JOURNAL      Unpublished (1986)
COMMENT      [2] revises [1].
            Draft entry and sequence in computer readable form for [2] kindly
            provided by A.O.Summers, 15-SEP-1986. Potential Shine-Dalgarno
            sequences are located at positions 157-162 (13.1 kD), 590-596
            (12.4 kD), 1740-1745 (59 kD) and 545-542 (15.9 kD). A single 38 bp
            inverted repeat, which delimits the leftward end of tn21, tn4 and
            tn501 extends from nucleotide 65-102. A promoter region for the
            structural genes is located between positions 366-585. This area
            contains regions of dyad symmetry which might function as operator
            sites for merR, which negatively regulates the expression of the
            structural genes as well as its own expression.

```





Db 1096 GCCGACTTGGCGGCATGAAACCGTGTCTGTGAATCGTGTGGCCATGCATGCACGA 1037  
 Oy 127 AAGGGGAATGTTTCCTCCCGCTTATCGGCTCAGTACAGG 167  
 Db 1036 AAGGGGAATGTTTCCTCCCGCTTATCGGCTCAGTACAGG 996  
 RESULT 10  
 AF071413 19672 bp DNA linear BCT 13-AUG-2002  
 LOCUS AF071413  
 DEFINITION Assembled sequence of transposon Tn21 containing transposition  
 genes, the integron In2, and the mercury resistance operon.  
 ACCESSION AF071413  
 VERSION AF071413.2 GI:21930298  
 KEYWORDS  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 REFERENCE 1 (bases 2952 to 4066)  
 AUTHORS Diver,W.P., Grinstead,J., Fritzinger,D.C., Brown,N.L.,  
 Altenbuchner,J., Rogowsky,P. and Schmitt,R.  
 TITLE DNA sequences of and complementation by the tnpr genes of Tn21,  
 Tn501 and Tn1721  
 JOURNAL Mol. Genet. 191 (2), 189-193 (1983)  
 MEDLINE 84013495  
 PUBMED 6312271  
 REFERENCE 2 (bases 18411 to 19672)  
 AUTHORS Mista,T.K., Brown,N.L., Fritzinger,D.C., Primmore,R.D.,  
 Barnes,W.M., Haberstroh,L. and Silver,S.  
 TITLE Mercury ion-resistance operons of plasmid R100 and transposon  
 Tn501: the beginning of the operon including the regulatory region  
 and the first two structural genes  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (19), 5975-5979 (1984)  
 MEDLINE 85014891  
 PUBMED 6091128  
 REFERENCE 3 (bases 15973 to 19672)  
 AUTHORS Barinau,P., Gilbert,P., Jackson,W.J., Jones,C.S., Summers,A.O.  
 and Wisdom,S.  
 TITLE The DNA sequence of the mercury resistance operon of the IncFII  
 plasmid NRI  
 JOURNAL J. Mol. Appl. Genet. 2 (6), 601-619 (1984)  
 MEDLINE 85159407  
 PUBMED 6530603  
 REFERENCE 4 (bases 16253 to 18549)  
 AUTHORS Mista,T.K., Brown,N.L., Haberstroh,L., Schmidt,A., Goddette,D. and  
 Silver,S.  
 TITLE Mercuric reductase structural genes from plasmid R100 and  
 transposon Tn501: functional domains of the enzyme  
 JOURNAL Gene 34 (2-3), 253-262 (1985)  
 MEDLINE 85322071  
 PUBMED 2989109  
 REFERENCE 5 (bases 2952 to 4066)  
 AUTHORS Rogowsky,P., Halford,S.E. and Schmitt,R.  
 TITLE Definition of three resolvase binding sites at the res loci of Tn21  
 and Tn1721  
 JOURNAL EMBO J. 4 (8), 2135-2141 (1985)  
 MEDLINE 86055774  
 PUBMED 2998784  
 REFERENCE 6 (bases 3057 to 4105)  
 AUTHORS Hyde,D.R. and Tu,C.P.  
 TITLE tnpm: a novel regulatory gene that enhances Tn21 transposition and  
 suppresses cotegrate resolution  
 JOURNAL Cell 42 (2), 629-638 (1985)  
 MEDLINE 85282620  
 PUBMED 2992807  
 REFERENCE 7 (bases 14935 to 15434)  
 AUTHORS Brown,N.L., Mista,T.K., Winnie,J.N., Schmidt,A., Seiff,M. and  
 Silver,S.  
 TITLE The nucleotide sequence of the mercuric resistance operons of  
 plasmid R100 and transposon Tn501: further evidence for mer genes  
 which enhance the activity of the mercuric ion detoxification  
 system  
 JOURNAL Mol. Gen. Genet. 202 (1), 143-151 (1986)  
 MEDLINE 86174347  
 PUBMED 3007931  
 REFERENCE 8 (bases 1 to 3056)  
 AUTHORS Ward,E. and Grinstead,J.  
 TITLE The nucleotide sequence of the tnpr gene of Tn21  
 JOURNAL Nucleic Acids Res. 15 (4), 1799-1806 (1987)  
 MEDLINE 87146495  
 PUBMED 3029727  
 REFERENCE 9 (bases 4106 to 7537)  
 AUTHORS Sundstrom,L., Radstrom,P., Swedberg,G. and Skold,O.  
 TITLE Site-specific recombination promotes linkage between trimethoprim-  
 and sulfonamide resistance genes. Sequence characterization of  
 dhfrv and sulI and a recombination active locus of Tn21  
 JOURNAL Mol. Gen. Genet. 213 (2-3), 191-201 (1988)  
 MEDLINE 89039710  
 PUBMED 3054482  
 REFERENCE 10 (bases 5384 to 5397; 6199 to 6253)  
 AUTHORS Schmidt,F.R., Nucken,E.J. and Henschke,R.B.  
 TITLE Structure and function of hot spots providing signals for  
 site-directed specific recombination and gene expression in Tn21  
 transposons  
 JOURNAL Mol. Microbiol. 3 (11), 1545-1555 (1989)  
 MEDLINE 90136084  
 PUBMED 2559298  
 REFERENCE 11 (bases 7540 to 9257)  
 AUTHORS Stokes,H.W. and Hall,R.M.  
 TITLE A novel family of potentially mobile DNA elements encoding  
 site-specific gene-integration functions: integrons  
 JOURNAL Mol. Microbiol. 3 (12), 1669-1683 (1989)  
 MEDLINE 90158115  
 PUBMED 2560119  
 REFERENCE 12 (bases 6301 to 6333)  
 AUTHORS Guerineau,F., Brooks,L. and Mullineaux,P.  
 TITLE Expression of the sulfonamide resistance gene from plasmid R46  
 JOURNAL Plasmid 23 (1), 35-41 (1990)  
 MEDLINE 90272799  
 PUBMED 2190244  
 REFERENCE 13 (bases 4106 to 5438)  
 AUTHORS Mercier,J., Lachapelle,J., Couture,F., Lafond,M., Vezina,G.,  
 Boissinot,M. and Levesque,R.C.  
 TITLE Structural and functional characterization of tnpl, a recombinase  
 locus in Tn21 and related beta-lactamase transposons  
 JOURNAL J. Bacteriol. 172 (7), 3745-3757 (1990)  
 MEDLINE 90299796  
 PUBMED 2163386  
 REFERENCE 14 (bases 14967 to 15674)  
 AUTHORS Allmeyer,H., Cresnar,B., Greck,M. and Schmitt,R.  
 TITLE Complete nucleotide sequence of Tn1721: gene organization and a  
 novel gene product with features of a chemotaxis protein  
 JOURNAL Gene 111 (1), 11-20 (1992)  
 MEDLINE 92192465  
 PUBMED 1312499  
 REFERENCE 15 (bases 19146 to 19199)  
 AUTHORS Summers,A.O.  
 TITLE Unwist and shout: a heavy metal-responsive transcriptional  
 regulator  
 JOURNAL J. Bacteriol. 174 (10), 3097-3101 (1992)  
 MEDLINE 92250400  
 PUBMED 1577681  
 REFERENCE 16 (bases 5134 to 5281)  
 AUTHORS Levesque,C., Brassard,S., Lapointe,J. and Roy,P.H.  
 TITLE Diversity and relative strength of tandem promoters for the  
 antibiotic-resistance genes of several integrons  
 JOURNAL Gene 142 (1), 49-54 (1994)  
 MEDLINE 94237488  
 PUBMED 8181756  
 REFERENCE 17 (bases 4492 to 8992)  
 AUTHORS Recchia,G.D., Stokes,H.W. and Hall,R.M.  
 TITLE Characterisation of specific and secondary recombination sites  
 recognised by the integron DNA integrase  
 JOURNAL Nucleic Acids Res. 22 (11), 2071-2078 (1994)

[illegible][illegible]

gene  
CDS  
complement(3003..3563)  
/gene="tmpr"  
/codon\_start=-1  
/transl\_table=11  
/product="putative resolvase"  
/protein\_id="AAG14403.1"  
/db\_xref="GI:10185691"  
/translation="MTGQRIGYIRVSTFPDNPERRQLEGVVDRAE SDKASGDVYRRL  
LEALISFARIGDITVVVHSMRLANLNDLRITVOTLQGVHIEFVKHLSFTGSDSS  
MALMLISVWGAEFEFELALREROREIDIALAKGAVGRKKSISSEIRIAELRQVEZ  
GQCKTLAEFEISRETLXYLRFDQ"  
complement(4242..5255)  
/gene="inti"  
complement(4242..5255)  
/gene="inti"  
/codon\_start=-1  
/transl\_table=11  
/product="putative integrase"  
/protein\_id="AAG14404.1"  
/db\_xref="GI:10185692"  
/translation="MKTATAPIDPLRSKVLDQLRERIRLYHSIPTEQAVVHVRAR  
IRFHVHRHATIGSSVEEAFISLANERKSVSTHRQALALAFYTKRVCTDILPWL  
EIGRRPSRRLPYVLPDPVRLITGLBGRHRLPAQLITGMRTISGLIDRLDLE  
DHGCTIIVRGKSKSRALMLPESLAPLSRLQSIRARAWMLKDAEGSVALPDALIE  
KYPKRGHSWPFVWFPAQHTHSTDPGRGVARNHMYDOTFORAKRAVEQAGITPATP  
HLRISFAALLRSGDIRTYDVLGHSDVTMIYTHVLKVGAGVRSPLDPLPIL  
SER"  
7600..8439  
/gene="sul1"  
7600..8439  
/gene="sul1"  
/function="may confer resistance to sulphonamide"  
/codon\_start=-1  
/transl\_table=11  
/product="dihydropteroate synthase"  
/protein\_id="AAG14405.1"  
/db\_xref="GI:10185693"  
/translation="WTVYFGILNLTEDSFDESRRLDPAGAVTAIEMLRVSGDVVY  
GPASHPARVYSPADELRIRLAPLIDALSDOMHRSIDSEFOPTQYALTKRGYLD  
IQGFDPPLPYDIAEDCLRIVYMSHAGRDGIATRGHLRPKDLDELIVFEFERSAL  
RRSSVAADRLILDPMGCFMSPAPETSLHVLNLKALGILPILSVSRKSEFAT  
VGLRVEDGPASLAELHAELHAGDGYVTHAPGDLSAITSFETLAKFNSRDKGL  
DHA"  
complement(14272..15177)  
/gene="mpba"  
complement(14272..15177)  
/gene="mpba"  
/function="may confer resistance to erythromycin"  
/codon\_start=-1  
/transl\_table=11  
/product="macrolide 2'-phosphotransferase"  
/protein\_id="AAG14406.1"  
/db\_xref="GI:10185694"  
/translation="MTVYTRADTSOLYLAARHGKLLHGPLTVNELGIDYRIVATVD  
DGRWYLIRTRRAEVSAAKEPEEARVLAHLKNRLPFAVPDWRVANAELVAYPMLEDS  
TA MYLQPGSTDPWVVDSEVFAEFAEFAALAAHLAHPISAADVAGMLIRPTPAROWA  
DYDVARREVFVNDKRLHRMQRWLDDDSWPDFSVYVGDLYGVHVLINTERVSGMI  
DYSEARDDAIDMAAHLWFGKEGLAKLLITVETAGGVWRPLAHLHIAERLAFGAVT  
YALFALDSGNEYEYLAAKAQAAAAE"  
16153..16968  
/gene="aph1"  
16153..16968  
/gene="aph1"  
/function="confers resistance to kanamycin"  
/codon\_start=-1  
/transl\_table=11  
/product="aminoglycoside 3'-phosphotransferase"  
/protein\_id="AAG14407.1"  
/db\_xref="GI:10185695"  
/translation="MSHIQRETSRPLNSNMADILGYKWARDNVGSGATIRLY  
GKPDAPLEFLKHKGSVANDVDEMVLNMLEMPLEPTIKHTRTPDDAULTTAIP  
GTAFVLYLEYPDSGENTYDALAVFLRLHSIPVCNCPFNSDRVRLAQAOSRMANGL  
VVASDDEDRNGVPEQVWKEMHKLLPFSPDSVYTHGDFSLDNLIFDGKLLIGCDVG

gene		RVGIDRYODIATLNNNGCEFSPLQKRLRFQKYGINDPMMNKLOFLMLDFFF"
CDS		/complement(21228..22088) /gene="bla" /complement(21228..22088) /gene="bla" /function="confers resistance to ampicillin" /codon_start=1 /transl_table=11 /product="beta-lactamase" /protein_id="AAG14408.1" /db_xref="GI:10185696" /translation="MSIQHFRVALIPFEAFCLPYFAHPDETIVKVADEQLGARVGY IELDLNSKILLESFPERPEPMSTFKYLLCGAVLSRVAGOGSLGRIRIHYSNDIVE YSPYTEKHLDGMYRELFCSAAITMSDNIAALULTTTIGSPKEITAFINMGCHYRL DRMPEELNEAIPINDERDTMPAAMATTIKILLTGELLTASRQQLIDMKEMADVVAGPL LRSLAPGMFIADSGAENERSKRGIIALALGPDKFSRIYVIYTTGSQATMDERRNOIA EIGSLIKHW"
repeat_region		complement(26615..26652) /note="Imperfect" /rpt_type=inverted
BASE COUNT	5647 a 7506 c 7633 g 5866 t	
ORIGIN		
Query Match	50.2%; Score 161; DB 1; Length 26652;	
Best Local Similarity	100.0%; Pred. No. 2.2e-71;	
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	7 CACTGCAGAGAGCCACAGCCTGTGGCCGAACAACAGTCAGAGAGTGGCGCAGAAGATG 66	
Dd	26425 CACTGCAGAGAGGCCACAGCCTGTGGCCGAACAACAGTCAGAGAGTGGCGCAGAAGATG 26484	
OY	67 GCCGACTTGGCGCGCATGAAAACCGTGCTGTCAACTCGTGTGGCCTGCATGCACAGA 126	
Dd	26485 GCCGACTTGGCGCGCATGAAAACCGTGCTGTCAACTCGTGTGGCCTGCATGCACAGA 26544	
OY	127 AAGGGAAATGTTCCCTGCCCCGTTGATGGCGCATACAGGG 167	
Dd	26545 AAGGGAAATGTTCCCTGCCCCGTTGATGGCGCATACAGGG 26585	
RESULT 12		
AP000342/c		
LOCUS	AP000342	94281 bp DNA linear BCT 22-FEB-2002
DEFINITION	Plasmid R100 DNA, complete sequence.	
ACCESSION	AP000342	
VERSION	AP000342.1	GI:5103148
KEYWORDS	merf; merrt; merp; merr; mera; merd; yada; yaea; tnla; tnibdelta1; yaha; ista; istab; ybdba; suli; qgcdebltal; aadal; intr; tnmp; tnpr; tnpa; ybja; cat; insa; insb; yqca; ycdb; stba; stbb; ycga; ycgb; ycha; ycia; ycib; ycja; yciB; ydaa; ydba; ydcb; ydca; ydda; ssb; ydea; ydeb; ydga; ydia; ydja; ydjB; yaaa; teer; tekr; tetr; tetc; teud; yeda; yerf; psia; psia; mok; nok; yeha; yeia; 3z; X; tram; finp; trau; traf; traB; traC; traC; traB; traB; traB; traB; traB; traB; traB; traB; yfia; yfib; yfiC; tirac; trbi; tiraw; traB; ygea; trbc; tran; traz; trfb; traf; trbA; trbQ; trbd; trbu; trbf; trah; trag; tras; trat; yhfa; trar; tral; trax; yleA; flno; yiga; yiGB; yiha; repA2; inc; repA6; repAl; repA4; tir; pemi; penk; yjca. Plasmid R100 (specific host:Shigella flexneri 2b strain 222, lab host:Pscherichia coli strain K-12) DNA. Plasmids. 1	
ORGANISM		
REFERENCE		
AUTHORS	Sampe1,G. and Mizobuchi,K.	
TITLE	Organization and diversification of plasmid genomes: complete nucleotide sequence of the R100 genome	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 94281)	
AUTHORS	Sampe1,G. and Mizobuchi,K.	
TITLE	Direct Submission	
JOURNAL	Submitted (18-May-1999) Gen-jichi Sampe1, The university of Electro-Communications, Department of Applied Physics and Chemistry; 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan (E-mail:sampe1@pc.nec.ac.jp, Tel:81-424-43-5492,	

```

gene
/translation="MSTKLTGTMTGTCSCAVHVKDLERKVPVQVSADSYANGSKAFIAI
EYCTSGDALLTAAVAGYRATPLADAPSVETPGGLDKMRDLGRNDKGTGSSGALIAV
TGGSGAAMAALIKAVBOGARVILLIERGTIGTCVANGCVSPKIMIRAIHILRESP
EDGIIATPITFORTALLAQQOARNDERLAEGLIENPAITVHIGSGAKDRNRL
IYVLNNGEVEVAVFDRCLLATGASPAVPIPGIKOTPYTSTEAIVSETIPKRLAVIG
SSVALLEIKOAFERLGAKVITILARSTLFRREDPAIGEATYAAPRHEGTEVRPHQAS
VAITIEGGESEFVLTTHAGELRADKLIVATGRAPNTRKIALDNTGTTLEPGQATIDP
GMETSEHIIYAADCCTDQPOFYVVAAGSTRAIINMGDALINTAMPAVFITDPQ
ATYGSIAEAEHHGCIKTDSTRTLIDVNPAPALNFDPRGIKIVLEGSGRLIGVAYA
PEGGELICORPALAIRRMVQYELADQLFPLTWEGELIKLAQTFNNKDVYKQLSCAG"
3459..3821
/gene="merd"
3459..3821
/gene="merd"
/note="99% identical to gp:R1004_7[merd of plasmid R100,
function unknown]"
/codon_start=1
/transl_table=11
/protein_id="BAA78790.1"
/db_xref="GI:5103154"
/translation="MSAYITVSQLAHNAGVSHIVRDLYVRGLLRPVACTTGCGYVDD
AALORLCFVPAAEFAGIGDALARLKRALDADAAGAAQALAVLRDLVERRRALHL
DAOLASPAERAEHEALP"
3818..4054
/gene="yada"
3818..4054
/gene="yada"
/note="100% identical to gp:ECOMERTET_3[urfl, putative
ORF, gene product of plasmid R100]"
/codon_start=1
/transl_table=11
/protein_id="BAA78791.1"
/db_xref="GI:5103155"
/translation="MNAAPDKLPETROPVSGYLGAVALVLCPLHPILAAVLATTA
GATGEHMGVAAALATGLFVAVTRILRAPFGS"
4051..4758
/gene="yaea"
4051..4758
/gene="yaea"
/note="99% identical to gp:ECOMERTET_4[urf2, putative ORF,
gene product of plasmid R100]"
/codon_start=1
/transl_table=11
/protein_id="BAA78792.1"
/db_xref="GI:5103156"
/translation="MTSSQPAQWTAAELDAQARQGLDLYHOPYLVDLRHRIAGAEAL
KMRAPRILGPPGOFLPLAESGILMEIGAMVLEGCACROMHKMGKQVPOFRIATINV
SASQVGPENDEKRVKLDMLPALELIEETSEYAFENPLAFSALRIAGVFAA
PDGFGYICNGLKCCPITTLTKIDGSPFARLPDARQDTIVRAVIOAHGIMDVIFRR
RRLHQLIRNGCCAASS"
4692..4716
/transposon="In2"
/gene="yaea"
/note="terminal IR of In2"
4833..6512
/gene="tnia"
4833..6512
/gene="tnia"
/note="100% identical to gp:TTW042226_8 [tnia of Tn21]"
/codon_start=1
/product="transposase of In2"
/protein_id="BAA78793.1"
/db_xref="GI:5103157"
/translation="MATDTPRIPEGCVATLPDEANERARRAETISPLAOSFVGEAE
ADMAALALISRQYVILIRARQSGVLTDVLPQOSGGGKRLPPEVRYVETHEL
OKFLIKRKSILAAAFHREVTQVCAKQKLEVPARNVALRIASLDPRKVIIRREGDAA
MYDVGEGEPATVATPEOVYDIDVITDLYVDNRDPIGRPYTLTIDVETFEVCVLA
RVLTLAPASVAGLCILVHACDKRMEGLNVEDMOMSGKPLLLYLDNAEFCVLE
LRGCGCHGRIDRYPRLGDPHYGCIYERIIIGTAMOKIHDELPGTFESNDQGDYDSL
NKRAITLRELERMVLAVGTIYHBSYHNGLLQPPARMMAVAVARVGPVAVTATSFLL

```





```

gene
/codon_start=1
/product="putative IS1 transposase"
/protein_id="CAD09622.1"
/db_xref="GI:16505744"
/translatation="MPGNRPYHGRWPDQHDFTSRKLRPGSVTERIQDPSGYICAKSR
OHMLFAYARLRTKVVAHVHGERTVMTALIGTMSLLSPFDVVIWMTDGMPLYESLKCK
LHVTSRYRORIERHNLIROHLRLARGRKLSFSVEIHDKIYGHLYLNIRKHQ"
complement(2265..2540)
/gene="insa"
/note="HCM1_05c"
complement(2265..2540)
CDS
/note="HCM1_05c"
complement(2265..2540)
/gene="insa"
/note="HCM1_05c, insA, probable IS1 transposase, len: 91
aa; highly similar to many from Enterobacteriaceae e.g.
SW:ISA1_ECOLI (EMBL:X52534), insA, Escherichia coli
Insertion element IS1 protein insA (91 aa), fasta scores;
E(): 0, 96.9% identity in 91 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop). Contains probable
helix-turn-helix motif at aa 67-88"
/codon_start=1
/transl_table=1
/product="putative IS1 transposase"
/protein_id="CAD09623.1"
/db_xref="GI:16505745"
/translatation="MASVSISCPSCSATGDGVNRGKSAGHQRYLCISCRKTWOQFT
YRASQPTGHOKIIDMANNVGCRATARMGVGLNLTILRHNLNSGRSR"
complement(2284..2289)
/misc_feature
/gene="insa"
/note="possible translational frameshift site, similar to
that determined experimentally (EMBL:X52534)"
complete(2472..2495)
/misc_feature
/gene="insa"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(2573..2585)
repeat_unit
/note="13 bp inverted repeat flanking IS1"
complement(2694..3644)
gene
/gene="corA"
/note="HCM1_07c"
complement(2694..3644)
CDS
/gene="corA"
/note="HCM1_07c, corA, probable magnesium and cobalt
transport protein, len: 316 aa; similar to many e.g.
SM:CORA_ECOLI (EMBL:L11042), corA, Escherichia coli
magnesium and cobalt transport protein (316 aa), fasta
scores; E(): 0, 99.1% identity in 316 aa overlap. Contains
Pfam match to entry PF01544 CorA, CorA-like Mg2+
transporter protein. Contains hydrophobic, possible
membrane-spanning regions"
/codon_start=1
/transl_table=1
/product="putative magnesium and cobalt transport protein"
/protein_id="CAD09624.1"
/db_xref="GI:16505746"
/translatation="MISAFQLNNRLTRLEVESQFLVNAAWIDLYEPDDDELRLOS
ELGOSIATPEPEDIEASARFFEDDDGLIHSEFFEDADHAGNSTVATTIDGRLF
VTRERLPAFLRYRMARSQSWMGNAVELLDLPETRIQLADEIENTYSIDEQSKR
VTMEHGQVLEDEVALSTLAIEDIGKWVRCLMDTORALNIVLRKRLPGEQANRE
ILRDISTLPHNESIFOKYNPLMQAMGFINIONTKIKFSVSVVFPLPIVASYS
GMNFEPDELKRREGYPAAIIPIITILAGLAPYLFRKKRW"
complement(2697..3584)
misc_feature
/gene="corA"
/note="Pfam match to entry PF01544 CorA, CorA-like Mg2+
transporter protein, score 432.90, E-value 2.8e-126"
complement(3653..3656)
RBS
/note="possible RBS"
4004..4007
RBS
/note="possible RBS"
4014..4343
gene
/gene="HCM1_08"
4014..4343
CDS
/note="HCM1_08, hypothetical protein, len: 109 aa; similar

```

to the N-terminal half of SW\_Y1GE\_ECOLI (EMBL:U02122),  
y1ge, Escherichia coli hypothetical protein (254 aa),  
fasta scores: E(): 0, 97.9% identity in 95 aa overlap"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09625.1"
/db_xref="GI:16505747"
/translacion="MAHQLLIGKMTTLNKRFLFATLLPFAVAADCAQSDPILT
VOAYIVNPQTERVKMTWOKRANGAWMTLHALLDVINSQGVCAAMNGILNLVFDIK
YRKKEVF"
complement(4363..4569)
/gene="HCM1.10c"
complement(4363..4569)
/gene="HCM1.10c"
/note="HCM1.10c, hypothetical protein, len: 68 aa; unknown
function"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09626.1"
/db_xref="GI:16505748"
/translacion="MGRWDTYTRLPQVLYDASRFHGCATAGQAGELTVAFSKPAID
DICKIVDTTAAEVLGKRISVAVES"
```

Query Match	50.2%	Score 161	DB 1	Length 218160
Best Local Similarity	100.0%	Pred. No. 1.6e-71		
Matches 161	Conservative	0	Mismatches 0	Indels 0
			Gaps	0
QY	7	CACATCGCGAGGGCGAGCAGCCCTGGCCCAACACAAAGCTTAAAGACGTGCGCGAAGATG	66	
Db	175130	CACATCGCGAGGGCGAGCAGCCCTGGCCCAACAAAGCTTAAAGACGTGCGCGAAGATG	175189	
QY	67	GCCGACCTTGGCGCGCATGAAACCGTGTCTGTGTGAACCTGCTGGCCCTGCCATGCACGA	126	
Db	175190	GCCGACCTTGGCGCGCATGAAACCGTGTGTGTGAACCTGTGTGCCCTTGCATGCACGA	175249	
QY	127	AAGGGGAATGTTTCTCTGCCCGTTGATCGGCTCACTACAGGG	167	
Db	175250	AAGGGGAATGTTTCTCTGCCCGTTGATCGGCTCACTACAGGG	175290	

FEATURES	source
REFERENCE	1 (bases 1 to 516)
AUTHORS	Osborn,A.M., Bruce,K.D., Strike,P. and Ritchie,D.A.
TITLE	Sequence Conservation between Regulatory Mercury Resistance Genes from Mercury Polluted and Pristine Environments
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 537)
AUTHORS	Osborn,A.M.
TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-1994) Osborn A. M., University of Liverpool, Genetics and Microbiology, Liverpool, Merseyside, UK, L69 3BX
FEATURES	location/Qualifiers
source	1..537

```
-10_signal      36..41
                  /citation-[1]
-35_signal      61..66
                  /citation-[1]
gene            72..527
                  /gene="merR"
CDS            72..527
                  /gene="merR"
                  /citation=[1]
                  /codon_start=1
                  /transl_table=11
                  /product="regulatory protein"
                  /protein_id="CAA83893.1"
                  /db_xref="GI:607037"
                  /db_xref="SPTREMBL:O57492"
                  /translation="MOINPENTLTIGVPAKAGVNETIRFYORKGLPEPDKPGYSIR
BASE COUNT      110 a 136 c 175 g 116 t
ORIGIN
Query Match      10.9%; Score 35; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 GTCCGCGAGAGATGCGCGACTTGCGCGCGCATGCA 86
Db 357 GTCCGCGAGAGATGCGCGACTTGCGCGCGCATGCA 391

RESULT 15
LOCUS ECU77087/c 1696 bp DNA linear BCT 14-MAR-1997
DEFINITION Escherichia coli plasmid R831b organomercury resistance (OMR)
            locus: mer operon regulatory protein (merR) and organomercurial
            lyase (merB) genes, complete cds.
ACCESSION U77087
VERSION U77087.1 GI:1679729
KEYWORDS merB; merR; OMR; organomercurial lyase; R831b; broad spectrum;
SOURCE mercurial resistance; plasmid.
ORGANISM Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE 1 (bases 1 to 1696)
AUTHORS Ogawa,H.I., Tolle,C.L. and Summers,A.O.
TITLE Physical and genetic map of the organomercury resistance (OMR) locus
JOURNAL Inorganic mercury resistance (Hgr) loci of the Incw plasmid R831b
MEDLINE 85155497
PUBMED 6099319
REFERENCE 2 (bases 1 to 1696)
AUTHORS Tolle,C., Tolls,P. and Summers,A.O.
TITLE Nucleotide sequence of the organomercury resistance (OMR) locus
JOURNAL Unpublished (1996)
REFERENCE 3 (bases 1 to 1696)
AUTHORS Liebert,C.A., Watson,A.L. and Summers,A.O.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-1996) Microbiology, University of Georgia, 527
            Biological Sciences, Athens, GA 30602-2605, USA
FEATURES
    source
        1..1696
            /organism="Escherichia coli"
            /db_xref="taxon:562"
            /plasmid="R831b"
            /complement(1..305)
            /note="similar to GenBank Accession Numbers M24940 and
            U60777; partial 5' tnpA gene and tnpA promoter region"
            /gene="merR"
            /complement(306..740)
            /gene="merR"
            /note="similar to GenBank Accession Number M24940; mer
```

```
operon regulatory protein"
/codon_start=1
/transl_table=11
/product="merR"
/protein_id="AAB49638.1"
/db_xref="GI:1679730"
/translation="MENNLENTLTIGVPAKAGVNETIRFYORKGLPEPDKPGYSIR
LAMEAVLSDIVACHSRQNSVCPLASLDQGTSLGASTA"
746..751
/note="possible merB -35 signal site"
751..768
/note="18 base pair hyphenated dyad; MerR binding site;
merO"
771..776
/note="possible merB -10 signal site"
complement(773..778)
/note="possible merR -10 signal site"
complement(793..798)
/note="possible merR -35 signal site"
811..1048
/note="similar to merA sequence, GenBank Accession
Numbers M15049 and K03089; truncated 3' nonfunctional end
of merA"
1049..1687
/gene="merB"
1049..1687
/gene="merB"
/note="similar to GenBank Accession Number M15049;
organomercurial lyase"
/codon_start=1
/transl_table=11
/product="merB"
/protein_id="AAB49639.1"
/db_xref="GI:1679731"
/translation="MKLAPYLIELLTSVNRNGTADLVLPLRELAQRVSRRTTLTG
ILDPAPRVAAVLEQATSTETDKDNITIGYGLTLETSYVFEDIRRLYAWCALDTLI
FPALIGRTARVSSHCAATGAPVSLVSPSEIDAVPAGAVSLVLPQEAADVROSPCC
HVHFASVPTAEDMAKSHQGLEGLAIVSHEAFVGEQERNRLQTMSSRTP"

BASE COUNT      382 a 487 c 460 g 367 t
ORIGIN
Query Match      10.9%; Score 35; DB 1; Length 1696;
Best Local Similarity 100.0%; Pred. No. 4.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 GGCACCCACTGCGAGAGCGCCAGCCTGCGCGCA 209
Db 506 GGCACCCACTGCGAGAGCGCCAGCCTGCGCGCA 472
```

Search completed: June 2, 2003, 08:35:36  
Job time : 1800 secs

